USAFA-TR-88-1

⑫

DTIC FILE COPY

PROCEEDINGS
PSYCHOLOGY IN THE DEPARTMENT OF DEFENSE

Ψ
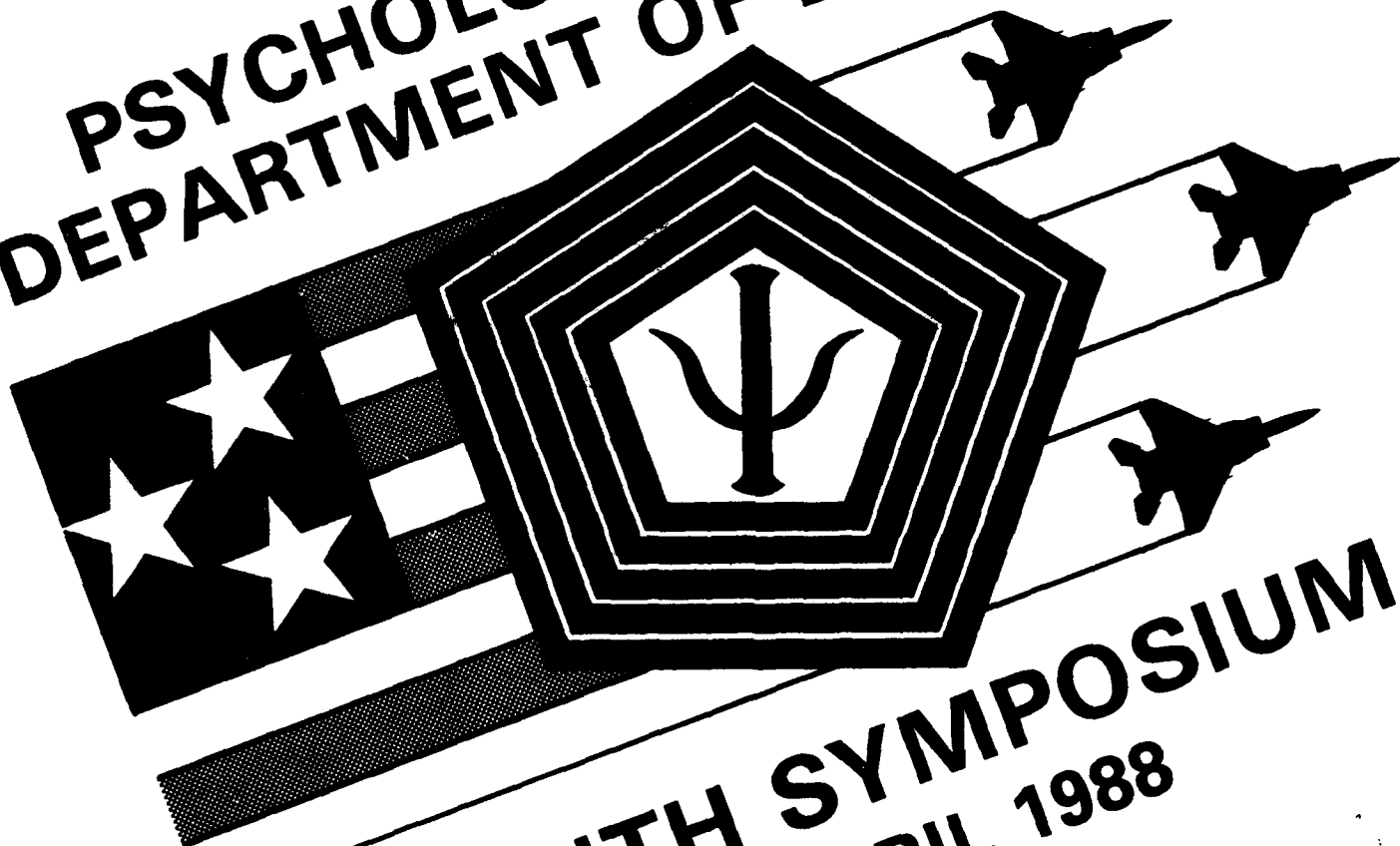
ELEVENTH SYMPOSIUM
13 - 15 APRIL 1988

DTIC
ELECTE
S
JUL 2 9 1988

A

SPONSOR:

USAF ACADEMY DEPARTMENT OF BEHAVIORAL SCIENCES AND LEADERSHIP

88

PROCEEDINGS OF THE ELEVENTH SYMPOSIUM
PSYCHOLOGY IN THE DEPARTMENT OF DEFENSE
13 - 15 APRIL 1988

DEPARTMENT OF BEHAVIORAL SCIENCES AND LEADERSHIP
UNITED STATES AIR FORCE ACADEMY
COLORADO SPRINGS, COLORADO

SYMPOSIUM CHAIR
LIEUTENANT COLONEL ROBERT C. GINNETT

PROFESSOR AND HEAD, DEPARTMENT OF
BEHAVIORAL SCIENCES AND LEADERSHIP

Colonel Richard L. Hughes

EDITOR

Major Francis E. McIntire

SYMPOSIUM COMMITTEES

PROGRAM COMMITTEE

Lt Colonel David B. Porter, Chair
Major Raoul J. Buron
Major Francis E. McIntire
Captain Gail A. Fallon

INVITATIONS AND REGISTRATION COMMITTEE

Captain Robert N. Shearer, Chair
Major Jeffrey S. Austin
Major Paul M. Grunzke
Captain Hugh A. Forde
Captain Thomas R. Mabry
Captain Andrew G. Stricker
1st Lieutenant Sharon L. Reinecke

SOCIAL, BILLETING, TRANSPORTATION COMMITEE

Captain Melissa R. Kallett, Chair
Captain Mark W. Patton
Captain Arthur E. Roffey

FACILITIES AND VISUAL AIDS

Captain Jeffrey A. Westberg

PROTOCOL
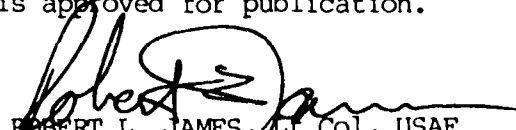
Captain Bruce M. Goeas

PUBLICITY

Captain Ann C. Miner

iii

This research report is presented as a competent treatment of the subject, worthy of publication. The United States Air Force Academy vouches for the quality of the research, without necessarily endorsing the opinions and conclusions of the authors.

This report has been cleared for open publication and/or public release by the appropriate Office of Information in accordance with AFR 190-17 and DODD 5230.0. There is no objection to unlimited distribution of this report to the public at large, or by DDC to the National Technical Information Service.

This research report has been reviewed and is approved for publication.

ROBERT L. JAMES, Lt Col, USAF
Director of Research, Studies
and Analysis

Additional copies of this document are avilable through the National Technical Information Service, U.S. Department of Commerce, 5285 Port Royal Road, Springfield, VA 22151. Telephone (804) 487-4660.

iv

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION	1b. RESTRICTIVE MARKINGS
UNCLASSIFIED	
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION/AVAILABILITY OF REPORT
	Approved for Public Release and Distribution Unlimited.
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	

4. PERFORMING ORGANIZATION REPORT NUMBER(S)	5. MONITORING ORGANIZATION REPORT NUMBER(S)
USAFA TR 88-1	

6a. NAME OF PERFORMING ORGANIZATION	6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION
Dept of Behavioral Sciences	DFBL	

6c. ADDRESS (City, State and ZIP Code)	7b. ADDRESS (City, State and ZIP Code)
U.S. Air Force Academy Colorado Springs CO 80840-5701	

8a. NAME OF FUNDING/SPONSORING ORGANIZATION	8b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
Air Force Office of Scientific Research		

8c. ADDRESS (City, State and ZIP Code)	10. SOURCE OF FUNDING NOS.			
	PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO.	WORK UNIT NO.

11. TITLE (Include Security Classification)
Proceedings, Psychology in the Department of Defense, Eleventh Annual Symposium (UNCLASSIFIED)

12. PERSONAL AUTHOR(S)
FRANCIS E. MCINTIRE, Major, USAF

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Yr., Mo., Day)	15. PAGE COUNT
Proceedings	FROM _____ TO _____	April 1988	528

16. SUPPLEMENTARY NOTATION

17.	COSATI CODES		18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB. GR.	Psychology, Symposium, Training, Leadership, Counseling, Cognitive Functioning, Operator Performance, Assessment. Stress, Job Performance, Turnover, Workload, Family Issues Biofeedback, Retention, Women's Issues, Learning. (Sixu)
05	05		
05	10		

19. ABSTRACT (Continue on reverse if necessary and identify by block number)
This printed proceedings include papers and presentation that deal with a wide range of research in psychology with emphasis on military issues. Keywords?

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT	21. ABSTRACT SECURITY CLASSIFICATION
UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS ☐	UNCLASSIFIED

22a. NAME OF RESPONSIBLE INDIVIDUAL	22b. TELEPHONE NUMBER (Include Area Code)	22c. OFFICE SYMBOL
FRANCIS E. MCINTIRE, Major, USAF	(719)472-2107	DFBLC

DD FORM 1473, 83 APR   EDITION OF 1 JAN 73 IS OBSOLETE

FOREWORD

The opinions expressed in this volume are those of the individual authors and do not necessarily reflect official policy either of the Army, Navy, Air Force, Marines, Coast Guard or of any government organization in which the author(s) may be serving.

Trade names of materials or products of commercial or non-government organizations are cited only where essential to precision in describing research procedures or evaluation of results. Their use does not constitute official indorsement or approval of the use of such commercial hardware or software.

These proceedings are approved for public release, distribution unlimited, in accordance with paragraph IVA, AFR 80-45.

In publishing the Proceedings of the Eleventh Symposium, the editor sought to facilitate an effective and timely dissemination of the technical information presented. The papers contained in this document were printed directly from unedited reproducible copy submitted by the authors who are soley responsible for their contents.

PROCEEDINGS OF THE ELEVENTH SYMPOSIUM ON PSYCHOLOGY
IN THE DEPARTMENT OF DEFENSE

MULTI-SERVICE STRESS PREVENTION AND INTERVENTION PROGRAMS

xxii

Simultaneous Performance of Rule-based and Knowledge-based Tasks

Donald Broadbent

A continuing problem in aviation, and in other fields, is the limit of unacceptable work load on the individual. Load can be reduced by separating the display of information into different sensory channels, by separating control actions into into different motor pathways, and so on. Speech annunciators and recognition systems now allow some expansion of these possibilities. However, an old experiment in the air illustrates the limits of this procedure. Many laboratory studies have since confirmed that it is not the whole answer. One particular group of recent experiments show a devastating effect of requiring verbal report of two near-simultaneous events, as opposed to requiring action in one task and inaction in another.

If complex decision tasks are simulated , those that are most vulnerable to a simultaneous task are those for which the person has reportable verbal knowledge. Tasks that are performed ´implicitly´ are less affected by a secondary task. This is true even though the inputs and outputs are of the same form in the implicit and explicit versions of the task. Thus it appears to be the distinction between explicit symbolisation and implicit, procedural, or rule-based function that alters the effects on work load.

FEATURED SPEAKER

DR. DAVID CAMPBELL
THE CENTER FOR CREATIVE LEADERSHIP

2

Automatic Information Processing
and Skill Acquisition

F. Thomas Eggemeier
University of Dayton

Abstract

Automatic processing theory provides a conceptual framework for skilled performance and guidelines for training to develop automatic components of such performance. General applications of automaticity principles to the development of skilled operator performance require extensions of current work in several important areas, including identification of the range of materials and conditions that permit automaticity, and specification of the transfer that can be expected with these materials under a variety of conditions. Recent research addressing these issues is reviewed, as are directions for future work.

The modification of performance that can result from extended practice has led a number of investigators (e.g., Logan, 1985; Schneider, Dumais, & Shiffrin, 1984; Shiffrin & Schneider, 1977) to suggest that substantial changes have occurred in information processing. Shiffrin and Schneider (1977), for example, have proposed a dual process theory which distinguishes between two qualitatively different forms of processing: controlled and automatic.

Within the dual process framework, automatic processing represents a rapid, parallel, and effortless process that is neither under subject control nor limited by short-term memory capacity. Controlled processing, on the other hand, is characterized as a relatively slow, effortful, and sequential process that is resource/capacity limited and controlled by the subject. Automatic processing develops with extended practice under consistently mapped (CM) conditions in which there is a consistent relationship between task components (e.g., a stimulus and the required response). Controlled processing is typically associated with either novel tasks or with variably mapped (VM) conditions in which task component relationships vary from situation to situation.

According to current dual process theory (e.g., Logan, 1985; Schneider et al., 1984; Shiffrin & Dumais, 1981), automatic processes represent essential components of many skilled behaviors. Important elements of skill develop from repetition and automatization of CM task elements, which results in the speed and efficiency that are characteristic of expert performance. Most skilled performance is, however, conceptualized as the product of both controlled and automatic processing (Logan, 1985; Schneider et al., 1984; Shiffrin & Dumais, 1981).

Because of their characteristic modes of operation (e.g., fast, relatively capacity-free), automatic processes are of great potential significance to skilled operator performance within complex systems. One such area is timesharing performance, since automatic processing of

selected task components could represent a means of reducing the loading
levels imposed by some current systems (e.g., aircraft, process control).
If the processing requirements of a task component were reduced through
automaticity, the freed capacity could possibly be applied to other
functions, thereby improving timesharing efficiency. Gains in timesharing
efficiency have been demonstrated in several instances in which automatized
tasks have been performed at high levels with concurrent controlled tasks
(e.g., Fisk & Schneider, 1983; Schneider & Fisk, 1984).

In addition to potential timesharing benefits, automatic processes
could improve operator performance in a number of ways. As noted above,
automatic tasks are performed more rapidly than controlled tasks (e.g.,
Fisk & Schneider, 1983; Hale, 1988). In addition, automatic processes can
result in more accurate (e.g., Myers & Fisk, 1987; Shiffrin & Schneider,
1977) and less variable (e.g., Myers & Fisk, 1987) performance than
controlled processes.

By stipulating that only CM elements of tasks can be automatized, dual
process theory also provides a prescription for training to permit
development of automatic components of performance. Since CM elements
represent important components of skilled behavior, training programs
should be structured to permit numerous repetitions and eventual
automatization of the CM elements of a task.

Automatic processing theory therefore provides both a conceptual
framework for skilled performance, and guidelines for training to develop
automatized components of skilled behavior. Applications of automatic
processing principles to train components that will exhibit the noted
advantages therefore have the potential to significantly enhance skilled
performance within complex systems.

General applications of automaticity to skill acquisition will,
however, require extensions of current laboratory work in several important
areas. Two particularly relevant areas include: (1) more extensive
specification of the range of task materials and conditions which permit
development of automatic processes; and (2) further investigation of the
transfer that can be expected with automatic processes for such materials
and conditions. This paper overviews selected research bearing on these
issues, and outlines directions for additional work.

Materials and Conditions Required for Automaticity

Complex systems (e.g., aircraft, air traffic control) are
characterized by the requirement to process a variety of materials (e.g.,
semantic, spatial) under situations that may involve consistency among only
particular components of the overall task, or which may be characterized by
so-called "higher-order" consistencies which are mediated by rules that are
not dependent upon the intertrial consistency of individual stimuli. Much
of the initial work with automatic processing (e.g., Schneider et al.,
1984) used alphanumeric stimuli in relatively simple search tasks in which
individual targets and distractors remained completely consistent
throughout the experiment. Although central to the development of dual
process theory, this work does not provide a strong basis for

4

generalizations of the theory to other important aspects of skilled performance.

Recent research, however, has addressed applications of automaticity to materials and conditions that are more representative of those found within complex systems. Work of this type has involved investigations of automatic processing in a number of areas, including: (a) semantic, spatial, and complex alphanumeric materials; and (b) conditions involving either higher-order consistencies, or consistent task components embedded in an overall task that incorporates inconsistent elements as well.

Applications to Complex Materials

A number of recent experiments (e.g., Eggemeier & Sullivan, 1986; Fisk & Schneider, 1983; Hale, 1988; Hassoun, 1987; Schneider & Fisk, 1984) have demonstrated evidence of automatic processing in either semantic category or word search paradigms. Hale (1988), for example, investigated the effects of extensive practice under both CM and VM conditions in a semantic category memory search task. At the completion of practice, CM conditions showed a significant advantage over VM conditions in the speed of the search process, and also demonstrated a significantly reduced slope in the function relating reaction time to the size of the memory search set. These differences are consistent with the development of automaticity in the CM condition, and suggest that automatic processing is applicable to the semantic materials represented in the search set.

Similar differences in CM and VM performance that are consistent with the automatization of semantic materials have been demonstrated by Fisk and Schneider (1983), Schneider & Fisk (1984), and Eggemeier and Sullivan (1986). These results extend the earlier work on automatic processing in alphanumeric tasks to an important type of material that must be processed by operators in a variety of complex systems.

A second important area for skill development within complex systems concerns the rapid and accurate processing of spatial pattern information. This type of processing is important for successful performance in systems such as aircraft or air traffic control, and a number of recent investigations have been successful in either: (1) demonstrating performance differences which are consistent with at least partial automatic processing of such materials (e.g., Eberts & Schneider, 1986); or (2) applying training principles derived from automatic processing to tasks tasks involving the processing of spatial information (Schneider, Vidulich, & Yeh, 1982).

Eberts and Schneider (1986), for instance, conducted several studies to determine the effects of extended CM and VM practice on the detection of line segment patterns which were presented sequentially on several channels of a visual display. A number of advantages of CM over VM conditions that are consistent with the development of at least partial automaticity were noted at the conclusion of practice. CM targets, for instance, were detected more reliably than were VM targets, and maintained that advantage when the number of channels to be monitored was increased.

An additional important application of automaticity is to the processing of complex alphanumeric materials. In some instances, operators of information systems (e.g., air traffic control) must search a display for alphanumeric patterns, and automatic processing of these materials represents an important issue in these applications.

Myers and Fisk (1987) have reported results from a series of experiments which support the applicability of automatic processing to such materials in a laboratory analog of a telecommunications industry job. Subjects were required to search fields in a display for patterns defined by the conjunctions of alphanumeric characters. After equivalent amounts of practice, CM patterns were detected more rapidly and accurately than were VM targets, and performance in the CM conditions remained invariant with increases in the size of the search set. These results therefore indicate that the benefits of automatic processing can accrue to complex alphanumeric materials.

Results of recent research with semantic, spatial, and complex alphanumeric materials therefore suggest that characteristics of automaticity that had been been identified with relatively simple alphanumeric materials do, in fact, generalize to materials that are more representative of those which must be processed by operators of complex systems. Some results, do, however, suggest that important limits may exist on the degree of automaticity that can be developed under certain situations. The Eberts and Schneider (1986) research, for example, raises important issues regarding the degree of automaticity that can be developed when target patterns are composed of sequentially presented spatial information, or when the stimuli to be processed are highly confusable. The capability to achieve only partial forms of automaticity under such conditions is of great potential importance in applications to complex systems, since these systems would typically impose such conditions on operator performance. Therefore, these possible limits should be more extensively explored in future work.

Applications to Higher-Order and Task Component Consistencies

As noted above, it is probable that important elements of complex tasks will involve higher-order consistencies, or consistent components embedded in an overall task. Recent research (e.g., Fisk, Oransky, & Skedsvold, 1988; Fisk & Schneider, 1984) has addressed the applicability of automatic processing to some such situations.

Fisk et al. (1988) investigated the role of higher-order consistency in the skill acquisition process. In one experiment, subjects determined which number in a display was the largest or smallest. Target and distractor numbers were drawn from the digits one through nine. Under the consistent decision condition, subjects made the same type of judgment on each trial, searching for either the highest or lowest number. In the inconsistent condition, the decision rule varied from trial to trail, requiring search for either the largest or smallest number. The consistent decision condition therefore incorporated a higher-order consistency, in that the same judgment rule was applied by subjects regardless of variations in the specific sets of digits that were presented from trial to trial.

Performance under the consistent rule differed markedly from inconsistent decision performance. Reaction times for consistent decisions were significantly faster than inconsistent decisions, and were invariant with increases in the size of the search set. Serial position effects related to the location of the target digit with respect to the end anchor of the set were also attenuated with practice in the consistent decision condition. In contrast, search set size and serial position significantly affected inconsistent decision times throughout practice. The noted differences in the speed of performance and in the search set size and serial position effects suggest that automaticity mediated by the higher order decision rule was developed with practice in the consistent decision condition. A second experiment demonstrated similar evidence of automaticity (i.e., attenuation of search set size and serial position effects) which was based on application of the same type of higher-order decision rule to a set of symbols (e.g., @, #) that included no pre-experimental ordinal relationships.

The results of the Fisk et al. (1988) work therefore support the position that local-level or stimulus-based consistency is not necessary for automatic process development if appropriate higher-order consistencies can be used by the subject.

In addition to higher-order consistencies, an important issue which bears on the applicability of automatic processing theory to skill acquisition in complex systems concerns the effect of embedding consistent task components in an overall task that also incorporates inconsistent elements. As noted above, current automaticity theory suggests that complex skilled performance represents a mixture of controlled and automatic processing, but the majority of laboratory work demonstrating the development of automaticity has been performed on relatively simple search tasks that exhibit consistent stimulus-response mapping throughout. However, if automatic processing theory is to be successfully applied to complex performance, benefits must be realized from consistent components that are embedded within the overall task structure.

The results of recent experimentation by Fisk and Schneider (1984) support the capability of subjects to utilize such component consistencies. This work employed a letter search task in which subjects indicated the position of a target in a four-element array by pushing a corresponding response button. Consistency in the detection/attention component of the task was manipulated by varying the relationship between the target and distractor sets. Under consistent attending conditions, particular letters remained either targets or distractors throughout, while under varied attending conditions, individual letters could appear as either targets or distractors across trials. Consistency in the response component was manipulated by varying the mapping between the position of the target and the appropriate response button. The mapping between spatial locations on the display and the response device remained constant under consistent response conditions, but changed across trials under varied mapping conditions. Consistent attending and consistent responding were factorialy combined across conditions in the experiment. The results indicated that consistent attending produced improvements in correct detections with practice, and further indicated that inconsistent responding failed to slow

the improvement rate. Inconsistent responding did reduce detection performance during training, but this reduction was eliminated when subjects were transferred to a completely consistent task.

The Fisk and Schneider (1984) results therefore suggest that automatic process development can contribute to total task efficiency by improving performance on consistent task components, even if the total task is not consistent.

The results of both the Fisk et al. (1988) and Fisk and Schneider (1984) experiments are important to applications of automaticity principles to complex skill acquisition, in that they extend the range of conditions for automatic processing development to instances that are representative of situations that are expected to occur in real-world systems. The Fisk et al. results significantly expand previous stimulus-based conceptualizations of the consistent task unit that can permit development of automaticity, and the Schneider and Fisk data demonstrate the feasibility of automatization of task components. Although both results represent important contributions, it is clear that additional work is required to build on these initial efforts and more fully explore the types of higher-order consistencies that can mediate automaticity, and the range of task component conditions which can facilitate performance. Such data would significantly contribute to the capability to apply automatic processing principles to complex skill acquisition.

Transfer of Automatic Processing

Another major issue pertaining to the application of automatic processing principles to complex skill acquisition concerns the transfer of automaticity that can be expected with the types of materials outlined above. Data concerning the conditions and limits of automatic component transfer with such materials would have important implications for the design of training programs intended to support the development of these components. Therefore, investigations of the influence on transfer of factors such as: (a) differing conditions of original training, and (b) varying degrees of similarity between the original and transfer materials are critical to eventual applications of automaticity to skill development.

A number of recent experiments (e.g., Eggemeier, Serafin, DaPolito, & Purcell, 1988; Eggemeier & Sullivan, 1986; Hassoun, 1986; Schneider & Fisk, 1984) have demonstrated transfer of automaticity with semantic materials of the type discussed above. Hassoun (1986), for instance, investigated transfer of automatic processing to unpracticed exemplars of previously trained semantic categories in both category and word memory search paradigms. Under category search conditions, the memory set (m-set) consisted of two semantic category labels (e.g., "Vehicles", "Fruits"), and subjects indicated whether or not a subsequently presented item (e.g., "Apple") was an exemplar of either category. After extensive CM training with a set of category exemplars, subjects were transferred to one of two search conditions: (a) Same Semantic Categories/Different Exemplars (S/D), or (b) Different Semantic Categories/Different Exemplars (D/D). At transfer, the S/D group demonstrated significantly faster reaction times than the D/D group, suggesting transfer to the new exemplars of the trained semantic categories.

8

Similar results were reported by Hassoun in a word search paradigm under high memory load conditions. During word search, the m-set consisted of exemplars of each of two semantic categories, and subjects indicated whether or not a subsequently presented probe item was a member of the set. Under a low memory load condition, the m-set included one exemplar of each category, while under high memory loading, three exemplars of each category were presented. At transfer, S/D performance was superior to D/D performance under the high memory load condition, but not under the low memory load condition. This result indicates the presence of potentially important transfer limits under word search conditions, and suggests that the locus of automatic processing and the degree of the resulting transfer may depend upon loading levels and the processing strategy adopted by a subject. Subjects might, for instance, choose to process semantic materials at the individual exemplar level under low memory load conditions, but at the more economical category level under higher levels of loading. Positive transfer in S/D conditions would result from the latter strategy, but not from the former.

Results which suggest other potentially important boundaries on the transfer of automatic semantic processing under category search conditions have been obtained recently by Eggemeier and Sullivan (1986). Eggemeier and Sullivan used a paradigm similar to that described above to investigate transfer of automaticity to exemplars of semantic categories that were similar in meaning to previously trained categories. Subjects were given extensive training under CM conditions in the memory search task, and were subsequently transferred to one of three conditions: (a) S/D; (b) D/D, or (c) R/D. The R/D condition consisted of exemplars of semantic categories that were related in meaning to the originally trained categories. As in the Hassoun (1986) work, the S/D condition demonstrated a significant advantage over the D/D condition at transfer. However, the R/D and D/D conditions did not significantly differ. These results indicate that under the conditions tested, automatic processing established with one set of semantic categories did not transfer to related categories.

The noted research demonstrating positive transfer of automatic processing within semantic categories is important in that it documents the occurrence of transfer with materials that must be processed by operators of complex systems. Both the Hassoun research on the effect of loading levels during original training and the Eggemeier and Sullivan work on transfer to similar materials suggest possible limitations on the transfer that can be obtained in the word and category search paradigms that were employed. It is clear that follow-on research is required to more extensively explore the boundary conditions of both results in order to document possible implications for transfer of semantic materials. At a more general level, future research should be conducted to address transfer with other types of materials (e.g., complex alphanumeric patterns) which are also of potential relevance to complex system performance. As was the case with the materials and conditions associated with the development of automatic processes, considerable additional work must be conducted to more fully explore conditions that affect the degree of transfer associated with automatic processes.

9

Summary and Conclusions

Current work in automatic information processing has extended the data base to several areas that are important to eventual applications of automaticity principles to skill acquisition among operators of complex systems. Within the context of the search paradigms that have been employed, this research has demonstrated the applicability of automatic processing to semantic, spatial, and complex alphanumeric materials. Similarly, recent work in higher-order and task component consistencies has indicated that automatic processing can develop under conditions that are representative of important constraints expected to be encountered in applications to real-world systems. Current research on transfer has demonstrated that automatic processing of semantic materials can be successfully applied to other instances of the same category, but has suggested some important possible limits on transfer that must be more fully addressed by future research.

In addition to work on the conditions and limits of transfer, future research aimed at facilitating the application of automatic processing principles to skill acquisition in complex real-world systems should more fully investigate the partial automaticity suggested by the Eberts and Schneider (1986) work with automatic processing of spatial materials, and should also address further refinements of the consistent trainable task component investigated by the Fisk et al. (1988) research on higher-order consistencies. This type of work should be accompanied by more extensive research on applications of automatic processing when only a component of the overall task is consistent. Such research is essential to a refined methodology for identifying those elements of actual operator tasks that can achieve some degree of automaticity, and should contribute to the successful application of automatic processing principles to complex skill development.

Acknowledgement

References

Eberts, R., and Schneider, W. (1986). Effects of perceptual training of sequenced line movements. Perception and Psychophysics, 39, 236-247.

Eggemeier, F.T., Serafin, C., DaPolito, K., and Purcell, S. (1988). Transfer of automatic category search. Unpublished manuscript, University of Dayton, Dayton, OH.

Eggemeier, F.T., and Sullivan, K.A. (1986). Transfer of automatic category search under conditions of varying degrees of similarity. Unpublished manuscript, Wright State University, Dayton, OH.

Fisk, A.D., Oransky, N.A., and Skedsvold, P.R. (1988). Examination of the role of "higher-order" consistency in skill development. Unpublished manuscript, Georgia Institute of Technology, Atlanta, GA.

Fisk, A.D., and Schneider, W. (1983). Category and word search: Generalizing search principles to complex processing. Journal of Experimental Psychology: Learning, Memory, and Cognition,10, 181-197.

Fisk, A.D., and Schneider, W. (1984). Consistent attending versus consistent responding in a visual search: Task versus component consistency in automatic processing development. Bulletin of the Psychonomic Society, 22, 330-332.

Hale, S.L. (1988). The locus of automatic processing in category word search. Unpublished master's thesis, Wright State University, Dayton, OH.

Hassoun, J.A. (1986). The effects of semantic feature similarity on automatic word search. Unpublished master's thesis, Wright State University, Dayton, OH.

Logan, G.D. (1985). Skill and automaticity: Relations, implications, and future directions. Canadian Journal of Psychology, 39, 367-386.

Myers, G.L., and Fisk, A.D. (1987). Application of automatic and controlled processing theory to industrial training: The value of consistent component training. Human Factors, 29, 255-268.

Schneider, W., Dumais, S.T., and Shiffrin, R.M. (1984). Automatic and control processing and attention. In R. Parasuraman and D.R. Davies (Eds.), Varieties of attention (pp. 1-27). New York: Academic Press.

Schneider, W., and Fisk, A.D. (1984). Automatic category search and its transfer. Journal of Experimental Psychology: Learning, Memory, and Condition, 10, 1-15.

Schneider, W., Vidulich, M., and Yeh, Y. (1982). Training spatial skills for air traffic control. In Proceedings of the Human Factors Society Twenty-Sixth Annual, 10-14.

Shiffrin, R.M., & Dumais, S.T. (1981). The development automatism. In J.R. Anderson (Ed.), Cognitive skills and their acquisition (pp. 111-140). Hillsdale, New Jersey: Erlbaum.

Shiffrin, R.M., and Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. Psychological Review, 84, 127-190.

FEATURED SPEAKER

DR. O.J. HARVEY
UNIVERSITY OF COLORADO

"Belief Systems and Some Educational Implications"

12

Asymmetric Transfer: Its Detection and Effect in Human Factors Research

Diane L. Damos, Ph.D.
Department of Human Factors
University of Southern California

Abstract

This paper discusses asymmetric transfer and its effect on human factors experiments. To demonstrate the magnitude of asymmetric transfer, the results of five experiments examining the effect of voice generation and recognition systems on dual-task performance are presented. The extent to which asymmetric transfer biased the data in three of the experiments is determined by statistical techniques and by comparing the data to the results of between-subjects experiments.

Asymmetric transfer, a carry-over effect (Kirk, 1968), occurs when the transfer from Condition A to Condition B is not the same as the transfer from Condition B to Condition A either in type (positive, negative, or zero) or quantity. For approximately 20 years, Poulton (1974, 1982; Poulton & Freeman, 1966) has warned human factors practitioners about asymmetric transfer and its effect on the results of within-subject designs. This warning has largely gone unheeded, perhaps because human factors practitioners have not realized the extent to which asymmetric transfer can bias data or how to detect it. This paper describes a series of five experiments examining the effect of voice generation and recognition systems on dual-task performance. The magnitude of the bias caused by asymmetric transfer is shown in several of these studies and the statistical indicants of asymmetric transfer are discussed.

The five experiments reported in this paper were part of a project begun in 1981. At this time six of seven experiments in the open literature demonstrated that dual-task performance was better when subjects responded to one of the tasks vocally rather than manually. Five of these seven experiments also investigated the use of voice generation systems. Four of these five studies showed an advantage for presenting stimuli for one of the tasks auditorily versus presenting visual stimuli for both tasks. All seven of these studies, however, used a tracking-discrete task combination. Consequently, to expand the existing data base, the experiments discussed in this paper were to examine the effect of voice generation and recognition systems on the performance of discrete task combinations. As the magnitude of asymmetric transfer in the initial experiments became apparent, however, the purpose of the experiments changed to include an examination of asymmetric transfer itself. Additionally, it

became necessary to re-examine the effect of voice generation and recognition systems on tracking-discrete task combinations.

The five experiments are discussed in the order in which they were executed. Because of space limitations, the author has not provided a detailed account of the procedure and results of each study. Instead, she has attempted to summarize the major results of each experiment, their implications, and the logic that led to the next experiment. The reader should, however, note the following: 1) Experiment 1 used different voice generation and recognition systems from those used in Experiments 2-5. 2) In all experiments the misrecognition rate was less than 1%. Because this rate was so low, the data were not corrected for misrecognitions. 3) All rejections were treated as correct responses during the experiment itself and for the subsequent statistical analyses. This treatment biased the data in favor of the speech response conditions. 4) Only statistically significant ($p < .05$) results of the dual-task analyses were reported. 5) Almost all of the analyses showed main effects of practice and interactions involving practice. These are not mentioned because they are not relevant to the major issues addressed by the experiments. 6) Performance on all tasks in Experiments 2-5 was measured using both speed and accuracy measures. The speed and accuracy scores were always examined for evidence of a trade-off. If a statistically significant trade-off was found, the data were analyzed using multivariate techniques. Otherwise, univariate techniques were used.

Experiment 1

The first experiment (Damos, 1985) had two major purposes. First, it examined the effect of a voice generation and a recognition system on the performance of a combination consisting of a verbal short-term memory task and a spatial short-term memory task. Second, it identified asymmetric transfer between stimulus modality combinations (auditory-visual versus visual-visual) of the two tasks. This experiment is presented in some detail to give the reader a general understanding of the procedures and the tasks used in all five experiments.

Method

For the spatial short-term memory task the subject saw 5 x 5 grids with five randomly illuminated cells. The subject's task was to determine if the current matrix was identical to the preceding matrix rotated 90 degrees to the left or right. If the current matrix was a rotated version of the preceding matrix, the subject pressed the "same" key, which was on a keypad under his left hand. If the matrix was different, the subject pressed the "different" key on the same keypad. As soon as the subject made a response, the matrix was erased and a new pattern presented.

14

For the verbal short-term memory task random digits between one and four were presented sequentially to the subject. The subject remembered the current digit and responded to the preceding digit. As soon as the subject made a response, the current digit was erased and a new one presented. Stimuli for this task could be presented either visually or auditorily. Subjects could respond either manually or vocally.

Sixteen male subjects were assigned at random to either a speech-manual response group or a manual-manual response group. Subjects in the manual-manual combination used their right hand to respond to the verbal task and their left hand to respond to the spatial (matrix) task; subjects in the speech-manual combination responded vocally to the verbal task and manually to the matrix task. Half of the subjects in each response combination group performed the verbal memory task initially with auditory stimuli and then with visual stimuli. The other half began with visual stimuli and then changed to auditory stimuli.

The experiment was conducted on four successive days. On Day 1 each subject performed 25 trials of the matrix task. On Day 2 the subject performed 20 trials of the verbal memory task using the response modality assigned to his group and the first stimulus modality. The subject then performed another 20 trials of the verbal memory task with the stimuli presented in the other modality. On the third day the subject received some practice on the matrix task and the verbal memory task with the stimuli presented in the first modality. The subjects then performed 25 dual-task trials. Day 4 followed the same procedure as Day 3 except that stimuli for the verbal task were presented in the second modality.

Results

An analysis of variance (ANOVA) conducted on the correct response interval (CRI--the time between correct responses ignoring any errors) of the matrix task showed a main effect of stimulus modality [F $(1,12)=12.84$, $p=.01$]. This effect indicates that subjects performed the matrix task better when auditory stimuli were used (2261 ms) for the verbal task than when visual stimuli (2573 ms) were used. There was also a Stimulus Modality X Order interaction [F $(1,12)=20.91$, $p<.01$]. This interaction indicates that asymmetric transfer occurred between stimulus combinations. An examination of the data revealed high positive transfer from the visual-visual to the auditory-visual combination. There was much less transfer, however, from the auditory-visual to the visual-visual combination.

The ANOVA performed on the CRIs of the verbal memory task showed a Stimulus Modality X Response Modality interaction [F $(1,12)=6.20$, $p=.02$]. This interaction indicates that subjects in

15

the manual-manual response groups performed better with auditory (1911 ms) than with visual stimuli (2431 ms). No difference occurred as a function of the stimulus modality when the subjects responded using speech. There was a significant Stimulus Modality X Order interaction [F (1,12)=5.01, p=.04). As in the analysis above, this interaction reflects asymmetric transfer between the stimulus combinations.

Because there is no way to correct for asymmetric transfer, any data biased by asymmetric transfer must be discarded to obtain a clear picture of the effects of the independent variables. Consequently, data from Day 4 were discarded and the ANOVAs recalculated using just the data from Day 3 (by eliminating the data from Day 4 , the design effectively became a between-subjects design rather than a mixed design). The ANOVA performed on the Day 3 matrix data showed no significant effects. The comparable ANOVA performed on the verbal memory data showed a Stimulus Modality X Response Modality interaction [F (1, 12)=4.91, p=.04). The interaction has the same interpretation given above.

Discussion

Three points need to be made about Experiment 1. First, the speech-manual combination did not result in better performance than the manual-manual combination. This finding is in direct contrast to the earlier studies discussed in the Introduction, the majority of which showed significant differences in performance favoring the speech-manual combination. Second, the ANOVAs performed on the verbal memory data and on the matrix data showed better performance when the verbal memory task used auditory rather than visual stimuli. These results are in accord with the results of the earlier experiments and with many human factors guidelines. Third, asymmetric transfer occurred between stimulus combinations. When the effects of asymmetric transfer were removed, all of the stimulus modality effects in the matrix data were eliminated. Further calculations indicated that the differences between the analyses performed on the data from Days 3 and 4 and the data from just Day 3 could not be explained in terms of statistical power.

Experiment 2

Because the results of Experiment 1 concerning the effects of voice recognition systems contradicted the earlier literature the P.I. decided to repeat the experiment (Damos & Lyall, 1986) with several methodological changes. The number of subjects was increased to 12 per group and the experiment was conducted on three consecutive days with dual-task conditions occurring on Days 2 and 3. The verbal memory task was changed from a simple memory task to a mental arithmetic task, the running difference task. This task required the subject to subtract the digit just presented from the preceding digit. As soon as the subject entered the difference, he received a new digit, which he

subtracted from the immediately preceding digit, etc. Again, the stimulus and response modalities of the running difference task were manipulated as independent variables. However, the design of Experiment 2 was changed so that the stimulus combination was the between-subjects factor and the response combination, the within-subject factor. Thus, the design used in Experiment 2 was the reverse of the one used in Experiment 1. This change permitted the identification of asymmetric transfer between response combinations.

Results

An ANOVA performed on the correct RT scores of the matrix task revealed a main effect of response modality [F (1, 20)=5.47, p=.03], demonstrating better performance when subjects responded to the running difference task vocally (2206 ms) rather than manually (2481 ms). The comparable ANOVA performed on the percent correct showed a main effect of response modality [F (1,20)=12.60], p<.01], indicating just the opposite: subjects performed better when they responded manually (85%) to the running difference task than vocally (81%). The statistical indicant of asymmetric transfer, the Response Modality X Order interaction, also was significant [F (1,20)=5.89, p=.02].

A multivariate analysis of variance (MANOVA) performed on the data from the running difference task showed a main effect of stimulus modality [F (2,19)=11.13. p<.01], revealing more accurate performance with auditory (90%) than with visual (75%) stimuli. Response modality also was significant [F(2,19)=7.83, p<.01], indicating more accurate performance when subjects responded manually (85%) than vocally (80%). There was also a Response Modality X Order interaction [F (2,19)=6.95, p<.01], indicating asymmetric transfer between response combinations.

To correct for the effects of asymmetric transfer found in the MANOVA and one of the ANOVAs performed on the matrix data, all the Day 3 data were discarded and the analyses were recalculated on just the Day 2 data. The ANOVAs performed on the correct RTs and the percent correct scores of the matrix task showed no significant effects. The MANOVA conducted on the running difference task data showed only an effect of stimulus modality [F (2,19)=15.74, p<.01], which can be attributed to differences in accuracy favoring auditory stimuli (89%) over visual stimuli (71%).

Discussion

The analyses performed on the data from this experiment showed three statistically significant response modality effects. One favored the speech-manual response combination over the manual-manual combination. All three effects vanished when the data were correct for asymmetric transfer, probably because of low statistical power. Again, on the running difference task,

17

performance was better with auditory stimuli than with visual stimuli.

Experiment 3

The magnitude of asymmetric transfer found in Experiments 1 and 2 was a cause of considerable concern; obviously, asymmetric transfer occurred both between stimulus combinations and response combinations. The size of the effect also was large enough to cause statistically spurious effects. Because the subjects in Experiments 1 and 2 experienced both of the within-subject factors under single-task conditions, Experiment 3 (Damos, 1986) used a strictly between-subjects design to avoid any possible bias caused by asymmetric transfer.

Method

Twenty-eight males completed the experiment. Each subject was randomly assigned to one of four stimulus/response combination groups. The experiment was conducted on two successive days. Day 1 was devoted primarily to single-task performance; Day 2, to dual-task performance. In all other respects, Experiment 3 was identical to Experiment 2.

Results

The ANOVA performed on the correct RTS of the matrix task showed no significant effects. The corresponding ANOVA performed on the percent correct scores showed a main effect of stimulus modality [F (1,24)=7.90, p<.01], favoring auditory stimuli (80%) over visual stimuli (72%).

The MANOVA performed on the running difference data showed a main effect of stimulus modality [F(2,23)=21.81, p<.01]. This effect is caused by more accurate but slower performance with auditory stimuli (89%, 2492 ms) than with visual stimuli (70%, 2000 ms). There was also a Stimulus Modality X Response Modality interaction [F(2, 23)=5.63, p=01], reflecting differences in accuracy between the four stimulus/response combination groups. The auditory-visual/manual-manual group had the most accurate performance (95%); the visual-visual/manual-manual group, the poorest (69%).

Discussion

Experiment 3 demonstrated again that the use of a voice recognition system did not improve performance on the running difference-matrix task combination although the use of a voice generation system did. The lack of significant effects favoring the use of a voice recognition system from Experiments 1-3 conflict with those of the earlier studies. Two explanations can be given for this conflict. First, asymmetric transfer biased the data of the earlier studies, resulting in statistically spurious effects favoring the use of voice recognition systems.

18

Second, voice recognition systems only improve dual-task performance in combinations with high response workloads, such as occur with tracking tasks.

To decide between these two explanations, Experiment 4 used the same task combination employed by several of the earlier studies, a one-dimensional compensatory tracking task with the running difference task. A within-subject design was used to allow asymmetric transfer between response combinations to be identified.

Experiment 4

Method

All stimuli for the running difference task were presented auditorily. The other task was a one-dimensional compensatory tracking task. The subject performed this task by making left-right movements of a control stick held in his left hand. The disturbance input to the cursor was a random forcing function composed of the sum of nine sine waves with a breakpoint frequency of 0.35 Hz. The transfer function was $Y=K(.5/s +.5/s^2)$. The dependent measure was the average absolute error expressed as a percentage of displayed scale. Thus, a score of 0% represented perfect performance (the cursor centered at all times during a trial) and a score of 100% represented the worst possible performance (the cursor displaced as far as possible throughout the trial).

Twenty-four males were assigned at random to one of two response order groups (speech responses for the running difference task first or manual responses first). The experiment was conducted on three consecutive days. Day 1 was devoted to single-task performance; Day 2, to dual-task performance with the first response modality. Day 3 was identical to Day 2 except that the subject performed the running difference task with the second response modality.

Results

The tracking data are shown in Figure 1. An ANOVA performed on these data showed a main effect of response modality [$F(1,22)=12.71$, $p<.01$], indicating that tracking performance was better when the subjects responded vocally (31% average absolute error) to the running difference task rather than manually (38% average absolute error).

A MANOVA performed on the running difference data revealed a main effect of response modality [$F(2,21)=12.65$, $p<.01$), reflecting faster and more accurate performance with manual responses (1575 ms, 91% correct) as compared to speech responses (1783 ms, 88% correct).

Although the Response Modality X Order interaction did not

19

reach significance in either analysis because of large individual differences in performance, large asymmetric transfer is apparent in Figure 1; there is no difference between the two groups on Day 2 but a statistically significant difference between them on Day 3. In this case the between-groups difference in performance on Day 3 is strictly a result of asymmetric transfer. Asymmetric transfer also was evident in both the percent correct scores and the correct RTs of the running difference task. On this task, however, asymmetric transfer acted to reduce the difference between the two groups on Day 3 rather than increase it, as it did with the tracking data.
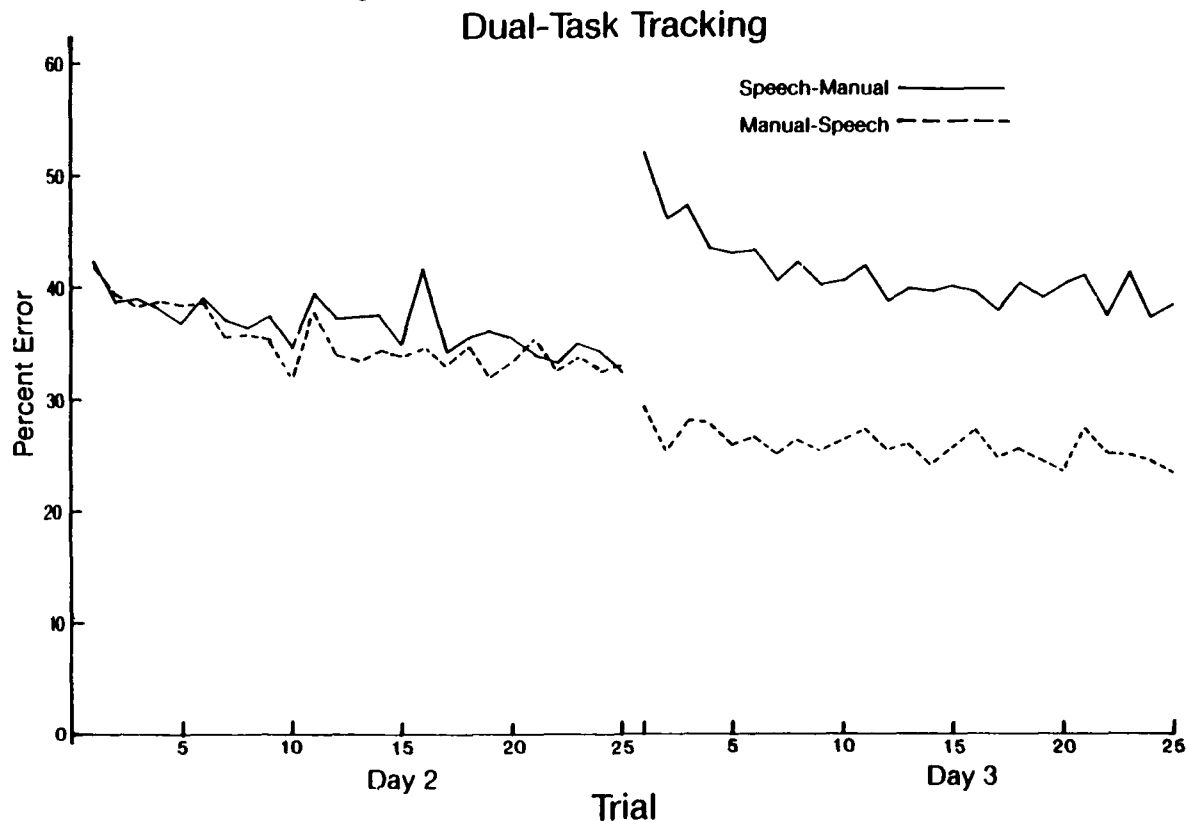
Figure 1. Dual-task tracking data from Experiment 4.

Discussion

The analysis performed on the tracking data from Days 2 and 3 indicated better performance when the subjects responded vocally to the running difference task rather than manually. This effect, however, was due entirely to asymmetric transfer; when the Day 3 data were eliminated, there was no difference between the response groups in performance. The MANOVA performed on the running difference data from Days 2 and 3 revealed significantly better performance with manual responses. The same effect was found when just the Day 2 data were analyzed.

This pattern of results indicates that, once asymmetric

20

transfer effects are taken into account, the use of a voice recognition system does not benefit dual-task performance in a tracking-running difference combination. This result contrad.cts most of the previous literature. Therefore, to provide some verification of these results, the experiment was "replicated" using a strictly between-subjects design (Damos, 1986).

Experiment 5

Method

The tasks were identical to those used in Experiment 4. The one major change from Experiment 4 was that the stimuli for the running difference task were presented either visually or. auditorily. Forty males were assigned at random to one of four stimulus/response combination groups. The experiment was conducted on two consecutive days. Day 1 was devoted to single-task performance; Day 2, to dual-task performance.

Results

The ANOVA performed on the tracking data showed no significant effects. The MANOVA performed on the running difference data showed a main effect of stimulus modality [F (2, 35)=5.04, p=.01], reflecting more accurate performance with auditory (91%) than visual stimuli (87%).

Discussion

Like the preceding experiments, this study showed significantly better running difference task performance when the stimuli were presented auditorily rather than visually. This study also confirmed the results of the other four experiment: the use of a voice recognition system did not improve dual-task performance, even on a combination with a relatively high manual workload. The use of a voice recognition system resulted in better performance than the use of a keypad only when the maximum recognition delay was subtracted from each speech response.

General Discussion

Two major points need to be discussed. The first concerns asymmetric transfer. This effect was found between stimulus conditions in Experiment 1 and between response conditions in Experiments 2 and 4. That is, every time the data were examined for asymmetric transfer, it was found. The simple presence of asymmetric transfer, however, is not the major problem demonstrated by these studies. The major problem, as discussed by Poulton (1982; Poulton & Freeman, 1966), is that asymmetric transfer can affect the data to such an extent that subsequent statistical analyses indicate spurious effects. The magnitude of asymmetric transfer is evident in Experiments 1, 2, and 4; all of these studies showed significant effects that were subsequently

eliminated when the data biased by asymmetric transfer were removed from the analysis.

The second major point concerns the advantages of voice generation and voice recognition systems. In all four of the experiments that examined auditory versus visual stimuli, auditory stimuli were found to result in better dual-task performance. These findings support much of the previous research on the use of these systems and are in accord with a number of human factors guidelines. The results of all five experiments provide little support for the use of voice recognition systems in the type of task combinations used in these studies. These results do not imply that voice recognition systems have no beneficial effect on dual-task performance; subsequent experiments by the author using a connected voice recognition system, a more advanced system than the ones used in these experiments, did show advantages for the use of the recognition system. The results of these five studies should, therefore, serve primarily to warn investigators of the magnitude of asymmetric transfer effects.

Acknowledgements

References

Damos, D. (1985). The effect of asymmetric transfer and speech technology on dual-task performance. Human Factors, 27, 409-421.

Damos, D. (1986). The effect of using voice generation and recognition systems on the performance of dual tasks. Ergonomics, 29, 1359-1370.

Damos, D. & Lyall, E. (1986). The effect of varying stimulus and response modes and asymmetric transfer on the dual-task performance of discrete tasks. Ergonomics, 29, 519-533.

Kirk, R. (1968). Experimental design: Procedures for the behavioral sciences. Belmont, CA: Brooks/Cole.

Poulton, E. (1974). Tracking skill and manual control. New York: Academic Press.

Poulton, E. (1982). Influential companions: Effects of one strategy on another in the within-subject designs of cognitive psychology. Psychological Bulletin, 91, 673-690.

Poulton, E. & Freeman, P. (1966). Unwanted asymmetrical transfer effects with balanced experimental designs. Psychological Bulletin, 66, 1-8.

TASKILLAN: A Validation Test Bed for Complex Performance Models

Christopher D. Wickens, John Flach, Arthur Kramer,
Kelly Harwood, and Gavan Lintern

University of Illinois
Institute of Aviation
Aviation Research Lab.

Abstract

In order to assess the validity of models of human learning and performance to real world aviation system design, a complex computer-based helicopter flight simulation was designed. This multi-task simulation, hosted on an IRIS visual graphics display system linked to an IBM AT, is described. Its relevance for the validity models of concurrent task interaction, of training, and of navigation and spatial cognition is discussed.

Introduction

Recent discussions in the human factors community have centered around the relevance, or lack of relevance, of basic laboratory research in engineering psychology to applied design problems in human factors (e.g., Meister, 1987; Simon (1987), Wickens (1987), (Kantowitz, 1987). As I have discussed elsewhere (Wickens, 1987), it is probably true that both sides-- experimental psychologists and human factors designers--must accept some portion of the blame for the hiatus between the two domains. Addressing the gap from the engineer's standpoint, it is necessary for design engineers to spend more time examining the basic theories that engineering psychologists have to offer, and assessing their potential relevance. Addressing the same gap from the perspective of the basic researcher, it is necessary to design and carry out experiments that test theoretical models at a level of complexity which is somewhat greater than the tight experimental control exerted in most laboratory studies. Only when performance models can be shown to be valid in these environments can they, or should they, be trusted by the design engineer.

The need for such validation is emphasized further by the very real potential for infusion of human performance models into the design process of complex military aviation systems. Examples of this potential are the Army's A[3]I program for advanced helicopter development (Corker et al., 1986), or the Air Force's cockpit automation technology (CAT) program (McDaniel, 1986). A major concern is that certain assumptions underlying many of the models that are being offered for such endeavors, have received little empirical validation. This is a particular characteristic of the assumptions that are made regarding the interaction between tasks in a multitask environment where models such as HOS (Harris et al., 1987; Siegel & Wolf, 1969; Laughery et al., 1986), PROCRU (Zacharias et al., 1981), WINDEX (North, 1985), or the task analysis methodology proposed for A[3]I (Corker et al., 1986), make assumptions regarding the degree to which different activities can or cannot be performed in parallel.

On the one hand, many of the assumptions are plausible and so it might be argued that any plausible model is better than no model at all. On the other hand however, the models differ radically in the degree to which they presuppose that parallel processing is possible. For example, PROCRU assumes that it is not. Laughery et al.'s adoptation of the Siegel and Wolf model assumes it is possible within the constraints of a limited capacity system. The model proposed by Corker et al. (1986) for the A^3I program assumes that parallel processing is not only possible, but will be perfect if demands are distributed across four structural channels. North's (1985) WINDEX model assumes computational interaction between channels, adopting a multiple resources approach (Wickens, 1984). We believe that the incremental value of these models will be greatly enhanced to the extent that these assumptions are critically tested in a complex task flight simulation.

The concern for evaluating models in complex task environments goes well beyond models of multiple task performance. The research program that we report, sponsored by NASA Ames Research Center, also addresses models of part-task training, of automaticity development, of spatial cognition, and of the effects of automation, although the latter two will not be described in this paper (but see Harwood, Barnett, & Wickens, in press).

A single helicopter simulation mission is used to evaluate the models, and the use of a single simulation is based upon an important rationale. That is, that it is too easy to configure specific simulations differently to validate each model, tuning the simulations either consciously or unconsciously in a way that will make the model "look good." The human factors practitioner, however, is not given that luxury when asked to apply his or her trade to system design. The system/mission requirements are specified in advance, and the applicability of the model must be demonstrated on a preexisting configuration of the system; not on one that is altered in a way to make application of the model most suitable.

The TASKILLAN Simulation

TASKILLAN, which stands for task skill analysis is a simulation, which uses an IRIS visual graphics system configured with an IBM AT, to represent a mission to navigate from a "home" to a target location where an operation is carried out, and then to return to the home base. The time-space representation of the mission is shown in Figure 1; static representations of the dynamic displays seen by the pilot are shown in Figure 2. The subject "flies" the simulator with a two-axis joystick that controls pitch and roll, along with a throttle that controls air speed. Altitude, and heading are the first derivations of pitch and roll, respectively, while in accordance with helicopter dynamics, air speed can be set to zero as the craft hovers. Flight parameters may be monitored on the HUD shown in Figure 2. As the mission is flown, the subject first traverses over a known path, and then enters an "unknown world" in which navigation is required from landmark to landmark: critical landmarks are indicated by voice, by printed designations, or by graphical paths or highlighting of objects represented on the AT-presented electronic map. After flying for 10-15 minutes through the unknown world, the pilot enters "Wall Street," a jumble of buildings, traversed at an altitude of 50 feet, well below the building tops. This environment simulates the absence of a distant spatial perspective, an absence that is characteristic of nap of the earth (NOE) helicopter flight. After emerging from Wall Street, the subject stabilizes the helicopter to perform the mission objective--tracking

Figure 1

Figure 2

and shooting at a series of evasive helicopters that traverse across the display using semi-predictable maneuvers. During this phase, the stability of the craft is reduced by introducing a positive feedback loop into the dynamics. The increased attention demands of the positive feedback instability make the flight control task much more difficult; but this demand may be eliminated by activation of an autopilot.

During the course of the mission, the primary task of flight control (stabilization and navigation), may be disrupted by the appearance of discrete side tasks. Some of these are computational, involving either arithmetic fuel calculations, or spatial target localization, while others involve the detection and classification of multidimensional threat stimuli. Still others involve schedulable "side tasks" or housekeeping chores. These side tasks may vary in the modality in which they are presented, or in their degree of demand (difficulty).

Models

Task Interaction: Multiple Resources. As noted above, current simulation models of complex task performance make different assumptions regarding the effect of task component load on total workload, the extent to which parallel processing is possible, and the degree of structural interactions between task channels (i.e., single vs. multiple resources, and the structuring of the latter). Using TASKILLAN, we are generating a wealth of data from operators in the time-sharing flight environment, from which we will be able to test the assumptions made by different models. Our particular interest focuses on three issues: (1) the extent to which workload drives the dynamic rescheduling of task components, (2) the best way of extracting component task demand estimates to predict complex task performance levels. We compare the use of performance measures, expert opinion, subjective rating, or algorithmic

25

models, (3) the degree of interaction between structurally different tasks; whether these are best predicted by a multiple resources approach (Wickens, 1984; Wickens & Liu, 1988; North, 1985), by a "parallel channels" approach (Corker et al., 1986), or by an undifferentiated capacity approach (Laughery et al., 1986). This issue is addressed in TASKILLAN by systematically varying the code of processing (spatial/verbal), modality of display (auditory-visual) of the side tasks, and by varying the demand levels of both the side tasks and the primary flight task--the latter as turbulence is introduced, and increasingly difficult phases of the mission are flown.

Models of Training: Part-task and Automaticity. In spite of years of research in the area, the issue of part-task vs. whole task training of complex skills remains far from resolved (Wightman & Lintern, 1985). Part of the lack of resolution stems from the absence of a theoretical basis or model underlying those transfer studies that have been carried out in complex environments, and the lack of complexity in those training studies that have been well-grounded in theory. In a review of the literature, Lintern and Wickens (1987) integrated the part-task training literature with attention theory to draw the following tentative conclusions: Heavy attention demand of component tasks will create an advantage for part-task training, but heavy time-sharing necessary for their integration will produce a reverse force favoring whole task training. Furthermore, if part-task training is undertaken, Lintern and Wickens concluded that task components containing perceptual and cognitive consistencies will benefit relatively more from the focus of attention allowed by part-task training than would task components that will not. Where such consistencies are absent, they concluded that a task component will be better served by training within the whole task configurations. The target acquisition phase of TASKILLAN in fact provides an ideal test bed for examining these predictions because the demands of what amounts to a four axis tracking task are quite extensive, while the two components (target acquisition and stabilization) do and do not, respectively, contain cognitive consistencies. Thus the theory predicts that target acquisition, but not stabilization, will benefit from part-task training. Our research results support the hypothesis of greater advantage for part task training of consistent tasks, but they also point to an overall advantage for whole task training in order to acquire the necessary time-sharing skills (Connelly et al., 1987)

A second theoretical approach to training, also tied to theories of attention, is automaticity theory (Schneider, 1985), a view which proposes that automaticity will develop to the extent that a task contains consistent mappings between its stimuli, or rules, and its responses. The TASKILLAN environment allows us to identify those components in which consistent mapping does take place and single them out for extensive part-task training, to determine if this procedure provides more efficient training than part-task training of varied mapping components. Our task analysis has identified threat identification as the logical candidate for consistent mapping, and our research program has pursued part-task training of this component task. In particular, we have examined the degree to which rules which may be consistently applied to categorizing targets as hostile threats, or friendly aircraft, can be trained to a level of automaticity, in order that those rules may be applied to target identification with minimal resource demand in the whole task simulation.

In summary, according to a task and skill analysis, the TASKILLAN

scenario provides a rich environment in which many of the basic principles and models of information processing should be manifest. But how important these principles may be, when surrounded by the other information processing complexities of the flight simulation remains an important question that must be addressed by empirical data. Only in this way, can the utility of those models to the human factors designer be realized.

Acknowledgments

References

Connelly, J.G., Jr., et al. (1987). Attention theory and training research. _Proceedings, 31st Meeting Human Factors Society_. Santa Monica, CA: Human Factors Society.

Corker, K., et al. (1986). _Development of an advanced task analysis methodology for Army-NASA aircrew/aircraft integration_. Bolt Beranek & Newman NASA Technical Report No. 6724.

Harris, R.M., et al. (1987). Microcomputer human operator simulator (HOS-IV). _Proceedings 31st Meeting Human Factors Society_. Santa Monica, CA: Human Factors Society.

Harwood, K., et al. (submitted). Situational awareness: A conceptual and methodological framework. _Proceedings, 1988 Psychology Symposium in the Department of Defense_.

Kantowitz, B.H. (1987). Training human factors psychologists. In J. Flach (Ed.), _Proceedings, 4th Mid-Central Ergonomics/Human Factors Conference_ (pp. 406-411). Savoy, IL: U. of Ill., Avia. Res. Lab.

Laughery, R., et al. (1986). A micro SAINT simulation analyzing operator workload. _Proceedings NAECON_. New York: IEEE.

Lintern, G., & Wickens, C.D. (1987). _Attention theory as a basis for training research_. U. of Ill. Inst. of Avia. Tech. Rep. (ARL-87-2/NASA-87-3).

McDaniel, J.W. (1986). Cockpit automation technology. _Contemporary Ergonomics 1986_. London: Taylor & Francis.

Meister, D. (1987). Reflections on the training of human factors specialists. In J. Flach (Ed.), (op. cit.).

North, R. (1985). WINDEX: A workload index for interactive crew station evaluation. _Proceedings NAECON_. New York: IEEE.

Schneider, W. (1985). Training high-performance skills: Fallacies and guidelines. _Human Factors, 27_, 285-300.

Siegel, A.I., & Wolf, J.J. (1969). _Man-machine simulation models_. New York: John Wiley & Sons, Inc.

Simon, C.W. (1987). Will egg sucking ever become a science? _Human Factors Bulletin, 30_, 1-4.

Wickens, C.D. (1984). _Engineering psychology and human performance_. Columbus, OH: Charles Merrill Pub. Co.

Wickens, C.D. (1987). Educating Human Factors Specialists: Some Important Concerns. In J. Flach (Ed.), (op. cit).

Wickens, C.D., & Liu, Y. (in press, 1988). Codes and modalities in multiple resources: A success and a qualification. _Human Factors_.

Wightman, D.C., & Lintern, G. (1985). Part-task training of tracking for manual control. _Human Factors, 27_(3) 267-283.

Zacharias, G.L., et al. (1981). _A supervisory control model of the AAA crew_ (BBN Report No. 4802). Cambridge, MA: Bolt Beranek & Newman, Inc.

LUNCHEON SPEAKER

DR. BONNIE DUNBAR
NASA ASTRONAUT

28

FEATURED SPEAKER

DR. CLAY FOUSHEE
NASA-AMES RESEARCH CENTER

Cockpit Crew Effectiveness from the Inside Out:[1]
A Micro-analysis Leading to Macro-considerations.

Lt Col Robert C. Ginnett, Ph.D.
United States Air Force Academy

Abstract

Crew failure is the largest single cause of aircraft accidents
yet the means of achieving effective crew work continues to be the
least understood of all aircraft systems. The research reported
here is the first phase of an ongoing program to investigate crew
performance as a group phenomenon. This micro-analysis of on-line
crews discovered four characteristics used by effective crew leaders
in the early developmental stages of their crews' lives. The
research also concluded that even at the micro-level, it is
necessary to consider the macro-factors if we are to understand
group work in organizational settings.

NTSB-AAR-73-14. In December 1972, Flight 401 crashed into the Everglades
killing 99 people. Investigation into the cause of the crash of this L-1011
revealed that there was only one malfunctioning electrical component--the nose
landing gear "down and locked" indicator bulb had burned out. The aircraft
was still flyable and landable. The gear was, in fact, extended and locked.
But the crew became so involved in trying to fix the one tiny indicator bulb
that they ignored the constant decrease in altitude that foreshadowed their
impact in the swamp. The crew failed.

NTSB-AAR-79-7. The crew of the Flight 173 had experienced only routine
conditions as they brought the four-engine DC-8 into the Portland, Oregon
traffic pattern in December, 1978. However, on final approach as they lowered
their gear for landing, they heard a dull thump from what seemed to be the
main gear area. (Actually, the main landing gear retract cylinder had become
disconnected.) Even though all systems indicated that the gear was down and
locked, the Captain elected to abort the landing and was put into a holding
pattern until they could determine if there was a problem and whether or not
it warranted further emergency precautions. This pattern was maintained for
approximately one hour at the Captain's insistence. During this time, both
the First Officer and the Flight Engineer warned the Captain on four separate
occasions that they were running out of fuel and needed to make a decision
about landing. In spite of these repeated cautions, the Captain (known within
the company as "Mr. DC-8,") insisted that they continue to circle. Finally,
as the first of the four engines flamed out, the Captain ordered the plane

---

toward the field while demanding that the Flight Engineer explain the cause of the engine failure. With all fuel tanks now dry, the other engines began to fail in sequence and the DC-8 nosed downward. About 1815 PST, Flight 173 crashed into a wooded, populated area killing 8 passengers and 2 crew members, and seriously injuring 21 passengers and 2 other crew members. Up until the time when the fuel was exhausted, the plane was flyable and landable, but the crew failed.

NTSB-AAR-82-8. "Slushy runway. Do you want me to do anything special for it or just go for it?" asked Roger Pettit, First Officer of Flight 90, as he peered into a snowstorm at Washington National Airport. "Unless you got anything special you'd like to do," quipped Larry Wheaton, the plane's 34-year old captain. Shortly after brake release, the first officer expressed concern with engine instrument readings or throttle setting. Four times during takeoff roll he remarked that something was "not right," but the captain took no action to reject the takeoff. (Air Florida operating procedures state the captain alone makes the decision to reject.) Seconds later, Flight 90 came back down, hitting the 14th Street Bridge before it crashed into the ice covered Potomac River, killing 74 persons on the aircraft and four motorists on the bridge. The NTSB ruled that the captain of the aircraft did not react to the copilot's repeated, subtle advisories that all was not normal during the takeoff. In spite of adverse weather conditions, subsequent computer simulations indicated that the plane was indeed flyable, but the crew failed.

NTSB/AAR-86/05. All the pilots and controllers in the Dallas-Ft Worth area on August 2, 1985 were aware of the thunderstorms and there were frequent requests for deviations around the buildups. Just after 1800 Flight 191 Heavy captured the ILS 17L localizer and glideslope on final approach, the First Officer (who was flying) said, "Lightning coming out of that one." "What?" asked the Captain. Again the First Officer repeated, "Lightning coming out of that one." "Where?" asked the Captain. "Right ahead of us," the First Officer replied. That was the last they talked about the severe weather directly in their path. The L-1011 proceeded into the buildup, encountered an unrecoverable microburst and crashed killing 134 on the aircraft and 1 person on the ground. Inspite of clear directives from a variety of sources prohibiting flight into thunderstorms, the crew failed to effectively perform their duties.

The anecdotal but increasingly compelling data from 1987's rash of airline accidents and incidents continues in this unfortunate trend. It continues to appear that not all crews operate with equal effectiveness. The second worst accident in U.S. aviation history occurred in Detroit on August 16, 1987. Although the National Transportation Safety Board has yet to release its final report, it appears that the crew, in a confused taxi sequence, failed to lower the flaps prior to takeoff.

The evidence is not merely anecdotal. The impact of ineffective crew performance was also demonstrated in a comprehensive review of jet transport accidents worldwide between 1968 and 1976 by Cooper, White, and Lauber (1979). They found that breakdowns in the performance of the crew as a whole played a significant causative role in more than 60 of the losses studied. And in a recent review of military transport aircraft accidents, the MAC IG reported that 80% were caused by crew error. Ruffell-Smith (1979) has empirically demonstrated the impact of crew failure in controlled simulator studies.

One argument frequently offered to explain the increasing number of accidents and incidents is the decreasing competence of the individuals flying today's airliners. While it is true that the relative experience level is going down, that does not appear to be a sufficient explanation. In the first

place, the individuals are still required to pass FAA examinations and check rides, both in simulators and on-line. There is no evidence that more pilots are failing or that the standards have changed. Secondly, if we had only one pilot in the cockpit, then it might make sense to look for an individually based causative explanation. But we do not have single-seat commercial aircraft--we have only crew-served aircraft. Therefore, the individual is only one component and perhaps not the most critical. If one member of the crew makes an error that is caught and corrected by another member of the crew, the crew has performed its duties. This, I would argue, is the bottom line.

Since it is a crew that matters in the ultimate measure of effective performance of commercial aircraft, and the evidence continues to show that crew failure causes accidents, it seems only reasonable to ask, why do we continue to look for individual solutions? A crew is a group--why not use a model of group effectiveness. Beyond that, it should be noted that cockpit crews are not intact work groups over time. It is very likely that any given dyad or triad comprising the crew on any particular trip has never worked together before (e.g. Foushee, 1984; Hackman and Helmreich, 1984). So if crews form frequently and some manage to perform quite effectively while others fail, what makes the difference?

Pilots report that whether or not a crew works well is in large part a function of the behavior of the captain. While they can not articulate the specific behaviors that an effective crew leader uses, they are able to identify effective and less effective captains with high reliability (Helmreich, Foushee, Benson, & Russini, in press). Furthermore, crew members consistently report that they can determine whether the captain they are about to fly with is effective (or not effective) at building and leading a crew very early in the group's life--"in just a couple of minutes." Whether this information is transferred in a formal "crew briefing" or informally in the first few moments in the cockpit, pilots claim that after only a few minutes with their captain, they will know enough about him or her to commence work as a crew. Given that most crews do not have a great deal of time for group development, this ability (if, in fact true) would serve a useful purpose. But if we accept this premise, we must question the ubiquitous claims of most "fixed sequence theories" of group development (e.g., Tuckman, 1965). Do crews really "form, storm, and norm" before they "perform?" Preliminary research into aircrew formation found little empirical data to support the notion that crews went through a period of "storming" enroute to performing.

The purpose of the first phase of this long term research project was to determine what happened in the formation process of airline flight crews, and to see if there were any differences between the behaviors of effective and less effective captains at building and maintaining their crews. (A final objective, to see if whatever happened in the early moments of the group's life carried over to line performance, will not be discussed in this paper.)

This research, which examined real work groups in an actual organization with a specific and consequential task to perform, used Hackman's (1986) normative model of work group effectiveness. This model also recognizes the importance of the first meeting. The steps outlined for "helping a team get off to a good start" seem particularly relevant for an assembling flight crew. These include (a) the development of boundaries, (b) coming to terms with the task, and (c) development of the norms to guide group behavior.

In addition to explicitly defining the boundaries, tasks, norms and roles, leaders (either formal or informal) can engage in behaviors that creates conditions such that the contributions of the other members are

33

encouraged and valued. In essence, leadership of the group is defined as functional behavior rather than in terms of traits or skills inherent in any one person. These behaviors, or the conditions which precipitate their emergence, may be expected to appear early in the life of the group. And while these functions might be performed by any member, at this point it seems reasonable to focus on the captain.

Hackman and Walton (1985) specifically examined the role of the legitimate leader in organizational work groups while maintaining a functional approach to leadership. This design draws explicitly from McGrath's (1962) definition that the leader's "main job is to do, or get done, whatever is not being adequately handled for group needs." One important area where a knowledgeable leader can help fulfill a number of critical functions is at the first meeting. In addition to the three functions described above, the effective leader might also take advantage of the opportunity to educate the group or to diagnose problems and opportunities the group might encounter.

Method: Phase I

Several factors dictated the methods for this research. As noted above, the models of group formation and development did not appear to fit the preliminary data on aircrew formation. Compounding this difficulty was the dearth of literature on the specific phenomenon of interest. A great deal had been written about crew performance--but nothing on crew formation or development. Finally, if the research was to be directly applicable to the crews whose formation it was attempting to explain, then it needed to be done in situ. As suggested by Hackman and Morris (1975), a controlled laboratory experiment might have eliminated the powerful variables underlying the phenomenon of interest. Therefore, this research was designed to be a field observation to collect and analyze largely qualitative data--a group level micro-analysis. All data were collected with crews which flew Boeing 727-200 aircraft within a single company.

Selection of Captains to be Observed

Captains who were particularly effective at building and managing their crew as a team and those who were conspicuously less successful in this regard were selected using a procedure involving nominations and evaluations by other pilots familiar with their cockpit behavior. Two check airmen involved in Line Oriented Flying Training (LOFT) development, crew effectiveness, and FAA coordination were asked to provide a list of names of captains who were extreme in one or the other direction on the above criteria. They were asked to nominate only those about whom they were personally knowledgeable, and to respond independently of one another. The two evaluators nominated a total of 37 different captains. Not unexpectedly, many names appeared on the lists of both nominators. These two lists were combined and sent to three other check airmen and the 727 chief pilot. They were asked to evaluate independently the names on the list, selecting twelve captains from each extreme. Their completed evaluations were sealed and mailed to an independent NASA researcher who maintained the confidential data base.

Captains selected for inclusion were those identified by the evaluators as among the most or least effective cockpit leaders. Operationally, this meant that a captain had been selected into his category by at least three of the four raters and that none of the raters had disagreed as to which extreme he should fall. Even if a captain received three nominations as an effective

34

crew leader, he would not be included in the pool if the fourth evaluator had nominated him as an ineffective captain in this regard. Of the 37 original submissions, 15 captains were selected as either high or low with "high concurrence." The final pool consisted of six effective crew management captains (labeled A Captains) and four captains who were less proficient as crew leaders (labeled B Captains).

Data Collection

The data were collected over the course of six months of crew and cockpit observation, including over 300 hours of direct crew observation and 110 hours of actual flight time. Prior to a scheduled trip, each captain was called in order to obtain his approval. In keeping with accepted research practices, each observed crew member was also asked for approval prior to in-cockpit data collection. Because the company in which the data were collected required a crew briefing one hour before the crew's first departure, intensive micro-level data collection (including tape recording, proxemic sketches, and notes) began at that time. On-line observations consisted of routine tracking of crew performance and intense data gathering during critical incidents for both task and process events.

The initial plan was for three trips to be scheduled with each captain. This was reduced to two trips later in the research after a lack of variation across trips was discovered.

<div align="center">Results</div>

Findings for All Captains

While an event actually occurs in the airline studied called the "Crew Briefing" and the entire crew is usually present for it, its content is not directed toward how the cockpit crew will work; it is directed instead to the cabin crew. Therefore, if one makes the a priori assumption that "the pilots" are "the crew," then the term "crew briefing" is a misnomer. In fact, the crew briefings have practically nothing in their content to do with cockpit crew performance, nor are the briefings directed to the cockpit members. On the other hand, the process of the crew briefing seems useful for getting an indication of the captain's, and consequently, the cockpit crew's performance. Even though the First Officer (FO) and Flight Engineer (FE) are usually present for the briefing, their role is generally limited to being introduced. This places them in the interesting position of actually being observers of the captain's briefing, thus providing them the opportunity to learn about his leadership behaviors.

One of the more surprising findings was the overall content and style similarity of briefings by each captain across time and crews. Quantitative analysis of the amount of time devoted to the crew briefing substantiated the captains' similarity across crews. The trend can be noted in the correlation between the time for the first and second briefings ($r = .73$).

The A Captains

Tasks. Of all elements in the normative model, this was the one least substantiated by the leader's behavior. There was virtually no mention of task work that was to be done solely by the flight attendants (i.e., checking seat belt compliance, serving beverages, etc.), nor was there any mention of

<div align="center">35</div>

task work for the group as a whole. All A Captains did, however, mention some aspect of task work that required interface between the cabin and the cockpit crew. Therefore, these tasks are, by definition, partly related to boundary issues. Directive statements made by captains included both the captain's task direction as well as the rationale for the decision.

Boundaries. Three boundary characteristics consistently emerged. The A Captains always entered the proxemic boundary established by the flight attendants, while (as noted above) the other cockpit crew members assumed positions to observe the captain's interactions. Second, all A Captains acknowledged the physical boundary that would separate the cabin crew from the cockpit crew on the aircraft (the cockpit door) and discussed how it should be managed. Third, the A Captains enlarged rather than restricted the relevant boundary for their work. They all included the cabin attendants as part of the relevant work team, and some expanded it further to include passengers, maintenance and ATC. They also included the researcher as an integral part of the crew.

Norms. Captains transmit normative statements both verbally and nonverbally. The A Captains made explicit statements about the way work should be conducted in one or more areas. These included safety, effective communication, and cooperation. Their nonverbal behaviors, both in the briefing and subsequently in the cockpit, supported their verbal statements.

Authority Dynamics. Even though not a part of the Normative Model, the management of the authority dynamic was observable for all A Captains. It must be noted that there is a heavy predisposition in all of aviation to establish and maintain the clear authority of the leader. Historically, there were obvious reasons for these relationships and legally, they still exist. But the problems with today's complex systems are not the same as in the early days of aviation. Time is not always the critical variable. Therefore, a leader who establishes and maintains appropriate conditions for crew work might behave quite differently than one who merely works to maintain the legitimate authority which is already his. In fact, that is precisely what the data indicated—but that is not all.

First, the A Captains established their competence in three ways. They gave logically organized briefings, they demonstrated technical competence about their job, and they were socially competent. But they offset their competence by demonstrating some imperfection (like lack of knowledge on some perfectly "knowable" matter) or some personal shortcoming. But perhaps most importantly, they "engage" their crew, not just in the subsequent critical work at the airplane, but in the briefing itself. Through "real-time" interactions, they involve the crew members as individuals in the work of the group. The effective leaders do not just brief their crews—they involve them in the briefing. Nowhere is the distinction between the A and B Captains clearer than here. It is not the case that the A Captains spent significantly more total time in the briefing with the crew ($\underline{M}$ = 388.83 sec.) than did their B counterparts ($\underline{M}$ = 259.87 sec.), $\underline{t}$=1.90, df=18, N.S. at the 0.05 level. Nor is it the case that the A Captains spent more time ($\underline{M}$ = 220.66 sec.) than the B Captains ($\underline{M}$ = 231.12 sec.) actually talking to the crew, $\underline{t}$=0.20, df=18, N.S. at the 0.05 level. There was, however, a significant difference in the amount of time that other members of the crew talked while the captain was present. The mean time for talking by others for the A Captains was 132 seconds/ briefing while the mean time for others talking during the B Captains briefing was 29 seconds ($\underline{t}$=3.089, df=18, $\underline{p}$ < 0.01). The process of teamwork does indeed begin early in the effective group's life.

36

The B Captains

None of the B Captains' overall behavior was similar to the patterns exhibited by the A Captains. Three of the four B Captains deviated from the A Captains in the early moments of the crew formation and their behavioral patterns remained consistent in subsequent crew performance. The fourth B Captain followed the general pattern of the A Captains in the crew briefing but deviated significantly from their behaviors in the cockpit. As discussed above, none of the B Captains exhibited the kinds of behaviors that actively engaged their crews. Other than that, the four B Captains deviated from each other nearly as much as they deviated from the A Captains. Inspite of these wide deviations, each B Captain failed to produce and maintain an effective cockpit team.

Captain Barry. This captain's behavior can be best summarized as "overcontrolling." In the briefing, his task instructions were not general but specific and not about what his behavior would be but about how others should do their tasks. Like the other B Captains, he did not engage the crew but demeaned them. His behavior was sexist and paternalistic toward the women occupying the roles of flight attendants.

Captain Brent. Captain Brent avoided conflict. The primary behavior under his control that had the potential to increase conflict was the exercise of authority--so if at all possible, he avoided this. Even when the exercise of his legitimate and expected authority was appropriate, he would not do it. In the crew briefing, he made few declarative statements and when he did, he "tagged" them which negated their impact.

Captain Bill. This captain had difficulty interacting socially with the individuals he was assigned to work with and this difficulty spilled over into other aspects of his capability to legitimately influence the group's life. He never became a part of the group with whom he would work--either spatially (in the briefing) or psychologically (at any time). His own definition of the purpose of the meeting held an hour before takeoff put him at a further disadvantage because of his lack of social skills. He repeatedly referred to the crew briefing as "the social hour," and questioned its utility. Given his conduct in the briefing, the utility of the meeting was indeed, questionable. His crew briefing, at its best, was short, and when it was expanded, it became more bizarre. Lacking any of the four positive categories of behavior seen with the A Captains, Captain Bill's briefing was a rambling dialogue of unrelated and tangential topics.

Captain Burt. Captain Burt was the only captain observed whose behavior in the early minutes of group formation was inconsistent with his categorization as a B Captain. The crew briefing was very much in the pattern of the A Captains, with the one exception being the lack of participative engagement of the other crew members as measured by time when they were speaking. However, his in-cockpit behavior clearly differentiated him from the A Captains as he lost control of his own behavior; hence he lost effective leadership of the crew.

Discussion

Clearly at the micro-level, the results demonstrated substantial differences between the behaviors of the A and B Captains in the early moments of the groups' lives and in subsequent line operations. Since reliable discrimination occurred in the eyes of experienced evaluators, it seems reasonable to conclude that there is such a thing as leadership and that

leaders can make a difference. As palatable as this conclusion may be to many, it is not a universally held tenet. For example, Pfeffer (1977) argued that the effects of leaders is small compared to the effects of other external or macro-variables. However, these crews were all operating under the <u>same</u> macro-variables (i.e., they were in the same organization, in the same industry, under the same regulations, etc.), yet the data show differences which can be attributed to the micro-behavior of the leader in the group setting. Perhaps Pfeffer's analysis focuses at the wrong level. At the same time, there are other questions which can best be answered by considering what Pfeffer might call the "macro-factors." It appears that the micro-analysis left several theoretical questions unanswered. The remainder of this paper is devoted to examining the importance of these macro-considerations for group effectiveness.

Let us begin this final section with a question. How is it that people who have never met can come together and in less than one hour, begin interdependent work without even talking much about it? (Recall that the least discussed element in the briefing was the task to be performed.) Critical to understanding this process is a preliminary understanding of the newly envisioned concept of the "shell." Like in computer science, a shell establishes a predefined sets of interactions that can occur between the computer and the operator thus making work easier.

The argument here is that a shell exists for the forming crew of a commercial airliner. The shell will be filled by a captain, a first officer, and a flight engineer and it will be linked to a team of flight attendants. This shell exists even before any of the individuals who will occupy the roles are assigned or are physically in place--it is a macro-variable. It provides a considerable amount of the appropriate role behaviors, norms, boundaries and authority dynamics that the crew will use in their work. The shell facilitates the interaction between sub-elements of the crew and even within other system elements. But these shell definitions are not rigid--they are more like definitions drawn with "dotted lines." They set expectations for behaviors by each member. It is this mechanism--the shell--which allows the crew to develop so quickly. When the captain begins his briefing, he is not really starting from a zero-base. If he is working in a company which encourages and supports crew effectiveness, then all of the contextual work which has gone before him in laying the groundwork is already in existence. All he has to do is merely <u>affirm</u> that imported shell--fill in the "dotted lines"--so the crew can proceed with that which they have come to expect. In these first few moments of the crew's life, the captain breathes life into the shell of the forming group. That only takes a few minutes but it is critical. The same "in-place shell" can explain precisely why the task is not discussed any more than it is. Every member of the crew is trained in their individual tasks and comes to work knowing what the work of the group is to be.

Captains can also do much more than affirm the shell, just as the A Captains did. They can <u>expand</u> the shell. The expanded shell, while providing the potential for more effective crew work, is nonetheless, quite fragile. In fact, the expanded shell is, itself, a new "dotted line" of how the group will work. Just as the "dotted line" of imported definitions for all forming crews requires some work by the leader, so too will the expanded shell require reinforcement. Some of that reinforcement is provided by the congruence between that which the captain "says" in the briefing and how he "behaves" in the briefing. Even more importantly is how he behaves during the flight. If he supports the expanded definition of crew work which he created, the "dotted lines" will fill in and more effective crew work will result. But if he

38

should behave in a manner inconsistent with the expanded shell, its newly formed and fragile outer boundaries will collapse. Effective team work doesn't just happen--it takes effort by everyone, but especially by the leader.

Unfortunately, not all captains have mastered the ability to expand or even affirm the shell of the forming group. These are the B Captains. Some of them simply abdicate their role leaving the "dotted line" in place with nothing added. The crew leaves the briefing not having learned anything about how their captain will behave. Even worse, some captains actively undermine the imported shell. These captains create the worst crews because the captain leaves them with no operable model for teamwork at all.

If we are to understand groups that work in the real world, then we need to pay more attention to the organizational context in which they work. But specifically, which organizational factors might have the greatest positive (or negative) impact on crew effectiveness? What macro-variables build the best shell for team work? Unlike the phenomenon labeled leadership, macro-variable do require a macro-analysis. That analysis is precisely the focus of the current phase of our research. Teams of observers from the United States Air Force Academy and Harvard University are currently studying cockpit crew work across organizational boundaries. Not surprisingly, some of the differences between commercial crew work and military crew work have already been determined. Our hope is to discover systemic differences in crew work that we can associate with variations in overall effectiveness according to the normative predictions.

References

Cooper, G. E., White, M. D., & Lauber, J. K. (Eds.). (1979). Resource management on the flight deck (NASA CP-2120). Moffett Field, CA: NASA Ames Research Center, (NTIS N80-22283).

Foushee, H. C. (1984). Dyads and triads at 35,000 feet: Factors affecting group process and aircrew performance. American Psychologist, 39, 885-893.

Hackman, J. R. (1986). The design of work teams. In J. W. Lorsch (Ed.). Handbook of organizational behavior. Englewood Cliffs, NJ: Prentice-Hall.

Hackman, J. R. & Helmreich R. (1984). Assessing the behavior and performance of teams in organizations: The case of air transport crews. Technical Report No. 5. Research Program on Group Effectiveness, Yale School of Organization and Management, Yale University.

Hackman, J. R. & Morris, C. G. (1975). Group tasks, group interaction process, and group performance effectiveness: A review and proposed integration. In L. Berkowitz (Ed.), Advances in experimental social psychology. New York: Academic Press.

Hackman, J. R. & Walton, R. E. (1985). The leadership of grops in organizations. Technical Report No. 6. Research Program on Group Effectiveness, Yale School of Organization and Management, Yale University.

Helmreich, R. L., Foushee, H. C., Benson, R. & Russini, R. (in press). Cockpit management attitudes: Exploring the attitude-performance linkage. Aviation, Space and Environmental Medicine.

McGrath, J. E. (1962). Leadership behavior: Some requirements for leadership training. Washington, DC: U.S. Civil Service Commission.

Pfeffer, J. (1977). The ambiguity of leadership. Academy of Management Review, 2:1, 104-112.

Ruffell-Smith, H. P. (1979). A simulator study of the interaction of pilot workload with errors, vigilance, and decisions. Moffett Field, CA: NASA-Ames Research Center, Report No. TM-78482.

Tuckman, B. W. (1965). Developmental sequence in small groups. Psychological Bulletin, 63, 384-399.

Organizational Behavior Before the Mast

Commander Earl H. Potter III, Ph.D.
Lieutenant William L. Zack, M.P.P.M.
U. S. Coast Guard Academy

Abstract

During the Fall Semester of 1987 the USCG Barque _Eagle_ sailed at the direction of President Reagan to Australia for Australia's Bicentennial celebrations. Embarked were 24 first class and 110 third class cadets in addition to 14 officers and a nucleus enlisted crew. As a result of extensive pre-cruise planning and a complete review of the impact of the cruise on academic programs, three courses normally offered as core courses at the Academy were taught in _Eagle_ underway. First class cadets took a 4.7 credit hour course in deck watch officer skills. Third class cadets received a two credit course in Oceanography and three credit course in Organizational Behavior. This presentation will deal with the experience of teaching Organizational Behavior (OB) on a sailing ship during a 13,000 mile voyage to the other side of the world.

Eagle departed New London, CT on 10 September 1987 with the echoes of speeches and fanfares. The OB course was focused on factors which contribute to a group's high performance and centered around a daily log kept by cadets in the course. In the log cadets recorded their daily experience, observations and reflections concerning their life and work in _Eagle_. The course readings, classes and films provided perspective on the cadets' experience. Cadets and instructors engaged in individual dialogues around cadet experience and their understanding of that experience as instructors read and commented bi-weekly on the logs and cadets responded to the prodding, suggestion and questions of the instructor.

The syllabus for the course was developed around the expected developmental cycle of the cadet watch sections (in brief, group formation, "storming and norming," performance in established teams, termination.) In general, we found this organization to work well; however, considerable adaptation was necessary. For example, we expected the period of group formation to begin when cadets reported to _Eagle_ a week prior to departure and last until we had been at sea for about two weeks. During this phase we expected to deal with issues of socialization, group membership, role definition, job design and a general overview of group development. However, soon after sailing we recognized a need to deal with issues of separation and grief. In order to look closely at where we were going we had to explore the impact of feelings cadets had about where they had been. Our second "surprise" was a realization that while the dominant theme in the first two weeks was a personal struggle to define each cadet's place in the organization, issues associated with later stages of group development, e.g., the formation of norms, identification of leaders and combinations of

cadets into pairs and cliques, were also relevant to the activity and performance of cadet work groups during this "phase." Perhaps the most striking realization for us as instructors was the great personal and team-based variety in ease and speed with which cadets dealt with the primary issues of group formations. For some, a week was sufficient; for others, three to four weeks seemed necessary.

In the second "phase" of our work groups' evaluation (storming and norming) we focused on issues of control, authority and motivation as well as the differentiation of watch sections by the development of "sub-cultures." Intruding into our plan were issues of stress and strategies cadets chose to deal with stress and confrontation with the culture of other countries. Arrival in Tahiti after a transit of the South Pacific under sail led to a dramatic shift in the "external environment." While some cadets began to think of going home, most became totally absorbed in he visits to six ports in four different countries within the span of 30 days. The change in focus had a major impact on group cohesion and performance which was more reflective of the way groups in practice evolve than the theory of group development. This phase of the course marked by increasing closeness among group members focused on issues of maintaining motivation and leadership in a now familiar organization. Finally, as Eagle cleared the kingdom of Tonga bound for Newcastle, Australia, cadets, while still required to perform, began to think seriously of going home. Maintaining high performance, dealing with group dissolution and the transition to a "new role" were the issues with which we closed the course.

It is difficult to tell whether student or teacher learned more in this course. As instructors we learned much about teaching OB with its implicit values in an organization which may not endorse those values. We also learned a great deal about the influence of our own values, wishes and fears on our teaching. Organizational Behavior, by its nature, encourages reflection on one's experience. We discovered that most reflective students—our "best" students—were not the cadets with a penchant for action who were seen as the most effective cadets. Since OB is taught in our curriculum with the expectation that a better understanding supports better performance, the observation that understanding and performance didn't "go hand in hand" posed some questions with which we are still wrestling. Furthermore, we became much more aware of the individual differences among our students which lead them to disconfirm and discount the wisdom of OB research which presents a statistically significant difference as a general principle. Finally, we were excited and rewarded by the impact of teaching OB as a dynamic "happening" within which we were involved and working to understand multiple transactions just as our students were.

Our presentation will begin with an overview of the planning that preceded the cruise, then move to a 30-minute video summary of the voyage and conclude with a discussion of the major learnings about OB, teaching and the place of OB in undergraduate military education.

Effects of an Intelligent CAI Tutoring System on the Spelling
Performance of Learning Disabled Students

Andrew G. Stricker
USAF Academy
Robert J. Hall
Texas A&M University

Abstract

Four pairs (2 male and 2 female) of learning-
disabled students matched for age, sex, aptitude, and
spelling achievement were instructed in spelling for
four weeks (45 minutes a day, five days a week) on two
versions of a computer-assisted instructional program
(SPELLDOWN). Treatments were alternated within pairs
and similar 10 word spelling lists were given to each
member of the pair. One version of the program
included elaborated correction subroutines featuring
scaffolded and faded cues customized to fit student's
individual responses. The control version of the
program was designed to mimic drill and practice
routines typical of classroom spelling instruction.
Analysis of the results may be interpreted to suggest
that the instructionally enhanced version of SPELLDOWN
produced improved quality in student's spelling errors
as well as increases in the number of words correctly
spelled relative to performance attained using the
control version of the program.

Microcomputer technology can be viewed as a potential
supplement to cognitive-behavioral training interventions to
assess and instruct students who have difficulty learning in a
specific academic domain (Hall, Gerber, & Stricker, 1988). Just
as early computers once provided psychologists with a new, albeit
limited, model of the components of human cognition, recent
developments have generated exciting new levels of
interdisciplinary exchange. On one hand, school psychologists,
for example, are beginning to seek understanding of computers as
potentially powerful tools for influencing students' ability to
profit from instruction. On the other, computer specialists and
engineers, interested in building "intelligent tutoring systems,"
are beginning to realize their need to have better models that
describe how real teachers and learners think. Moreover, through
computerized formative assessment of learners, it may be possible
to gain insight into the strategic and tactical decisions used by
individuals within and across problem-solving environments
(Wenger, 1987).

This study investigated the impact of an intelligent
computer-assisted instructional (ICAI) tutoring system on the
spelling knowledge of students with learning problems.
Intelligent tutoring, as opposed to "computer assisted

instruction," refers to a teaching system designed to communicate a body of knowledge, not merely to administer and monitor opportunities to practice simple responses. Knowledge, in the sense it is used here, means possession of specific facts, but also possession of information about how to use those facts. Spelling knowledge was selected to be the instructional domain in this study not merely because it includes a set of sound-symbol translation rules, or a corpus of words one can spell, but also a set of procedures, tactics, and strategies for using such information in the service of written communication. Normally achieving students were not selected for this study because teaching those students who bring significant knowledge and skill, preorganized as it were, to be optimally receptive to available modes of instruction, is not difficult. In fact, these students could be characterized by their reliable ability to demonstrate achievement gains in response to short verbal explanations, abbreviated demonstrations and models, outlines of general strategies, disclosure of useful rules and algorithms, or simple instructions or inducements to perform. While the success of these students is usually taken as evidence of "good" teaching, we should frankly admit that the observed outcomes are at least as much the result of a "good" student. That is, normally achieving students possess knowledge to be built upon, but also exhibit knowledge about knowledge -- what it is, why one needs it, who has it, how to acquire it, when it's sufficient for some specific purpose, and so forth.

Part of what makes them "normally achieving," is their ability to use their well-integrated knowledge of language, not only for speaking, reading and writing, but also for communicating effectively with teachers about speaking, reading, and writing. They have and easily use the capacity to recognize uncertainty, ambiguity, insufficient and irrelevant information, and error. Moreover, they also have the ability to prompt for additional data, to frame questions, and to "understand" even relatively abstract explanations. Thus, when teaching these students, teachers can communicate only, we might say, on one channel. That is, teachers need only to transmit basic information in the form of general explanations because the teacher's language maps onto an internal language and knowledge base that normally achieving students already possess.

Students experiencing learning difficulty, however, require teaching communications on two channels, one that transmits the elemental facts, and the other which models a structure for relating and using those facts. That is to say, teachers must make effective thinking in a domain transparent so that students can observe the thinking as well as its content. Use of this "second channel" of teaching communication is inherently difficult for teachers because it is cognitively demanding and time consuming. Therefore, students that require such teaching communication are rightly viewed as difficult to teach. It is important to point out that students experiencing learning

44

difficulties are within the normal range of intelligence and are not just children. Research in the area of learning disabilities has been interpreted to suggest that underdeveloped strategic abilities to select, organize, and coordinate cognitive resources in the presence of demanding but solvable academic problems may best characterize the difficulties encountered by learning-disabled students (Gerber & Hall, 1987; Hall, 1980; Koligian & Sternberg, 1987; Swanson, 1987).

Spelling is an interesting academic task to use for investigating the effectiveness of ICAI to help learning disabled students because (1) spelling products are permanent, thus, amenable to detailed review, (2) spelling attempts are generated in the absence of all but acoustic constraints, and (3) spelling attempts require interactive use of bottom-up and top-down processing problem-solving strategies (Gerber, Hall, & Stricker, 1988). In teaching spelling, for example, this requires not only communicating a letter sequence or a rule for generating a letter sequence, but also it requires teaching language that explicitly probes what internal understanding the student possesses about spelling that will permit the student to remember and appropriately use knowledge of specific letter sequences or spelling rules. Moreover, based on iteratively derived models of the learner's current status, this second channel of teaching communication must also consist of a teaching language that is deliberately and economically constructed as it is being used to model for students a way of thinking and talking about logical or empirical relationships that exist among different spelling sequences or rules.

Thus, intelligent tutoring, whether incorporated in hardware and software, or embodied in professionals, is quite complex. Beyond expert knowledge, it requires the ability to communicate that knowledge in an orderly and applied fashion when confronted by domain-related problems. In addition, the knowledge to be communicated must include standards for regulating and evaluating its use so that students internalize how, not just what, to learn.

A program called SPELLDOWN currently under development by Hall and Stricker (1988) and Stricker and Hall (1988) has attempted to incorporate expert knowledge on spelling and to communicate that knowledge in an orderly and applied fashion when confronted by unique learning problems inherent in a given student. SPELLDOWN was designed as a five-day, self-contained tutorial system for teaching and evaluating performance in spelling. The five-day package corresponds to the amount of time allocated in most classrooms for presenting, practicing, learning, and testing lists of words. General formats for the five-day program are as follows:

Day 1 - presentation of tape-recorded words and corresponding sentences;

Day 2 - tachistoscopic presentation of words;

Day 3 - presentation of sentences with spelling words
missing;

Day 4 - competitive spelling (child vs. computer) to
foster development of spelling fluency; and

Day 5 - standard spelling test format (e.g., word, word
used in sentence, word)

On days one and three, error prediction and detection data
are collected for each word. In addition, the program is
designed to preserve for each spelling variant, (1) inter-letter
latencies, (2) time to initial response, (3) total time, and (4)
letter sequence. For words spelled incorrectly, scaffolded
problem-solving cues are introduced to (1) guide student's
responses; (2) provide a general model for problem solving in
spelling; (3) provide extended practice in how to correctly spell
words; and (4) teach phoneme-grapheme correspondence rules
through systematic introduction of phonetically similar, regualr
words. Design features such as branching and faded cueing are
used to customize feedback for students needing more elaborated
constraint-seeking information. Speed and accuracy data
collected by the computer are stored in a series of relational
data bases linked via algorithms. These mathematical, decision-
making formulas help the computer to determine the type and
extent of assistance needed by students. Correct and incorrect
response feedback, level of help available for any given word,
and length of tachistoscopic presentation are determined by the
teacher at the time spelling words are entered into the program.
Figure 1 depicts the structure and relationships between the
various components of SPELLDOWN.

The first thing that one might notice about SPELLDOWN is
that the program changes from day to day. That is, there are
slight variations in the program format on different days. We
feel that instructional programs need to have identifiable
consistency so that students can predict what will happen next
and thus become comfortable with the learning environment.
However, it is important to vary, in systematic ways, how
information is presented to and requested from students. This
tends to lessen the effects of context dependent learning and to
increase the probability that a broader range of cues will be
available to students faced with spelling problems. From the
direct instruction literature, we have adopted the notion that
errors must be corrected immediately and that in correcting
errors, the process of producing a correct variant should be
modeled for the child. SPELLDOWN provides students with
opportunities to view their own work and to compare their
spelling attempts against correct standards. In that way,
students can immediately see precisely if and where errors have
occurred. The program tracks students' problem-solving

46

Figure 1. SPELLDOWN system structure.

EXPERTISE MODULE

1. Practice routine [condition-action pairs]:
   a. Copy the spelling word
      1. If first attempt then repeat, give messages 20,21
      2. If second or more attempt then repeat, give messages 22,21
   b. Unscramble spelling word with special help
      1. If misspelled then rescramble, give message 24
   c. Spell the word from memory with special help
      1. If misspelled then repeat, give messages 25,21
2. Special help routine structure [condition-action pairs]:
   a. If first attempt on letter:
      1. If letter not in word then flash word, give message 27
      2. Letter in word but in wrong position, give message 28
      3. If vowel is required and wrong vowel is selected, give message 32
   b. If second attempt on same letter:
      1. If vowel then display vowels in word, give message 33
      2. If consonant then display consonants in word, give message 34
   c. If third attempt on same letter then give correct letter and lock student in at word position, give message 29
3. Self-monitoring structure using messages: 2,3,5,6,10-14,17,22,25,27-34,39
4. Specialized spelling lists featuring graphemic features

INTERFACES

1. Student:
   INPUT   mouse with onscreen special keyboard
   OUTPUT   message boxes
2. Teacher:
   INPUT   keyboard and mouse through system administration program
   OUTPUT   student performance files and reports
3. Data files interfaces:
   Temporary buffers - holds current on-task performance data

   Permanent storage - holds previous performance data

   Current performance variables are moved to permanent storage when 50 [k] variables are accumulated.

SPELLDOWN
SYSTEM

TUTORIAL MODULE

1. Teacher specified constraints:
   a. SHE = 1 No special help for list
       = 2 Special help during practice
       = 3 Always provide special help
   b. H[k] = 1 Never provide special help
        = 2 Autosetting: provide special help for word during practice
   c. CT3 = Flash speed setting for word list
2. Teacher specified feedback for correct and incorrect spellings
3. Teacher specified messages to student at start of tutorial
4. Teacher specified student-database search depth [DB]
5. System generated homework assignment based on tutorial performance

Tachistoscopic presentation speeds:
   .5 seconds
   1.0 seconds
   1.5 seconds

STUDENT MODEL

1. Previous performance variables from student database:
   a. ERO2 - error rate
   b. REL2 - spelling reliability
   c. PER2 - percentage correct
   d. WRO2 - total wrong
2. Current on-task performance variables
   a. ZS[k] - selected spelling variant
   b. TIME[k] - spelling elaspe time in seconds
   c. LAT[GN] - current inter-letter latency time in msec
   d. LETTERS[GN] - current letter
   e. CCR - current word position in list
   f. C - total correct words
3. Current on-task learning behavior variables:
   a. R[k] - self-monitored recognition of difficulty before spelling attempt
   b. R2[k] - self-monitored recognition of difficulty after spelling attempt
   c. SPH - depth of special help
   d. WDIS[k] - spontaneous self-corrections for variants

47

strategies by monitoring the critical information-processing indicators of uncertainty and speed, order, and type of response. The instructional scaffolding provided by the program adapts to the student's progress on the basis of those indicators and previous performance on words containing similar problematic phonetic features.

Method

The Tandy Corporation sponsored the research to assess the impact of SPELLDOWN on the spelling performance of students with learning problems. Tandy provided four Tandy 1000SX computers with 20MB hard cards, high resolution color monitors, mouse controllers, hardware interfaces, and printers for this study. Each student work station also had a portable audio cassette tape recorder, controlled by the computer, and headphones. Computer work stations were set up in a local elementary school classroom.

Dependent variables. On days one and three, error prediction and detection were collected for each word. In addition, the program preserved each spelling variant, inter-letter latencies, time to initial response, total time, and letter sequence.

Subjects. Participants for the study were drawn from a pool of forty elementary school children participating in a summer Research and Diagnostic Teaching Clinic operated by the Department of Educational Psychology in conjunction with the Generic Special Education program at Texas A&M University. All students were enrolled or had been referred for placement in special education resource rooms at their respective schools. All Clinic children were administered the brief form of the Kaufman Test of Educational Achievement, the Test of Written Spelling, and the Spellmaster Assessment, Regular Word Test. Using age, sex, and performance indicators, four matched pairs of children (two male and two female) were selected to participate in the study.

Procedure. The study included two levels of analysis: by individual and by matched pair. The latter was included to control for effects of age, sex, and general aptitude. Two treatment conditions were also included: SPELLDOWN and a modified version of SPELLDOWN, which did not include the extensive modeling, feedback, or instructional features of the parent program. The lobotomized version of SPELLDOWN was computer-based and patterned after the parent program such that it served as a control condition for each of the subjects. On a week-by-week basis for four weeks, students alternated between SPELLDOWN and the control program, although each student in the respective pair worked on the same 10 word spelling list during the week. Typically, children spent 45 minutes working on the daily tutorials. Each week, a new 10 word list, derived from the Spellmaster Assessment and Teaching System, was introduced to

each pair of students. In sum, each student was exposed to 40 regular spelling words during the four week period that data was collected. Students spent two weeks working on words presented using the drill and practice configuration of the control program and two weeks using the constraint seeking cues and prompts provided by SPELLDOWN.

Results

Analyses of the data, may be interpreted to indicate that students learned to spell difficult words more rapidly and completely when guided by the SPELLDOWN program than when participating in the control drill and practice program. These differences held when students were compared to themselves or to their matched peers. Accuracy measures included bigram analysis, spelling quality ratings, and percentage of correct spellings. Of note, was that students using the modified computer program did demonstrate increases in performance. However, increases were greater following participation in the SPELLDOWN program.

Regression analysis using proportion of correct bigrams as the dependent variable yielded a significant treatment effect favoring SPELLDOWN performance. This analysis compared weekly individual records for each matched pair across treatments. The magnitude of the effect averaged 31% higher along the continuum of each tutorial day for students using SPELLDOWN (see Table 1).

Table 1

Treatment Effect Between SPELLDOWN and LOBO on Bigram Accuracy Proportions

	Week		
Pair	2	3	4
1	.21708	NA	NA
2	.5787	*	NA
3	*	.2724	.2012
4	.27058	*	.34806

* = intercepts not significant

NA = data incomplete

Finally, a one factor repeated measure analysis for proportion of correctly spelled words over treatments indicated

that SPELLDOWN was significantly more effective than the control
program (F [1, 10] = 21.5, p < .0009), see Table 2. It would
appear that the intelligent CAI spelling tutorial system
(SPELLDOWN) resulted in improved quality of student's spelling
errors as well as increases in the number of words correctly
spelled relative to performance attained using more traditional
drill and practice methods.

Table 2

One Factor Repeated Measure Analyses For Students on Proportion
of Correctly Spelled Words Over Treatments

Source	DF	SS	MS	F	P
MODEL	11	.85658571	.07787143	3.55	p < .0279
ERROR	10	.21958341	.02195834		
TOTAL	21	1.07616912			
SUBJ	10	.38447481		1.75	p > .1953
TREAT	1	.47211090		21.50	p < .0009

Treatment	Mean
SPELLDOWN (A)	.67118
LOBO (B)	.37820

    Figures 2 and 3 illustrate observed treatment effects, using
bigram accuracy proportions, for one pair of students over a two
week period. Bigrams are defined as consecutive letter sequences
and equal the total number of letters in a word plus one. For
example, the word "big" has four bigrams, _b - bi - ig - g_.
Blanks are included at the beginning and end of words to control
for letters added by students in initial or final positions. As
can be seen in each graph, Brad and Jeff were accurate in
representing about 50% of the total number of bigrams present in
the 10 word list on the first day. We think that this finding is
important for two reasons. First, correctly representing 50%
percent of all possible bigrams on first exposure is a good
indicator that the words chosen were difficult but not
frustrating for these students. Neither student refused to
attempt any of the words. Second, both students achieved the
same level of performance on the first day regardless of the
treatment condition. We intrepret this to mean that we have
identified an appropriate match. Words were equally difficult
for both boys on initial exposure. Looking across days 2, 3, 4,
and 5, however, it can be seen that performance differs
dramatically for the two students. Using the SPELLDOWN program,
Brad by day 2 during week 4 and Jeff by day 3 during week 5,

Figure 2. Bigram accuracy proportions for week 4.

Bigram Accuracy Proportions - Week 4/Pair 5
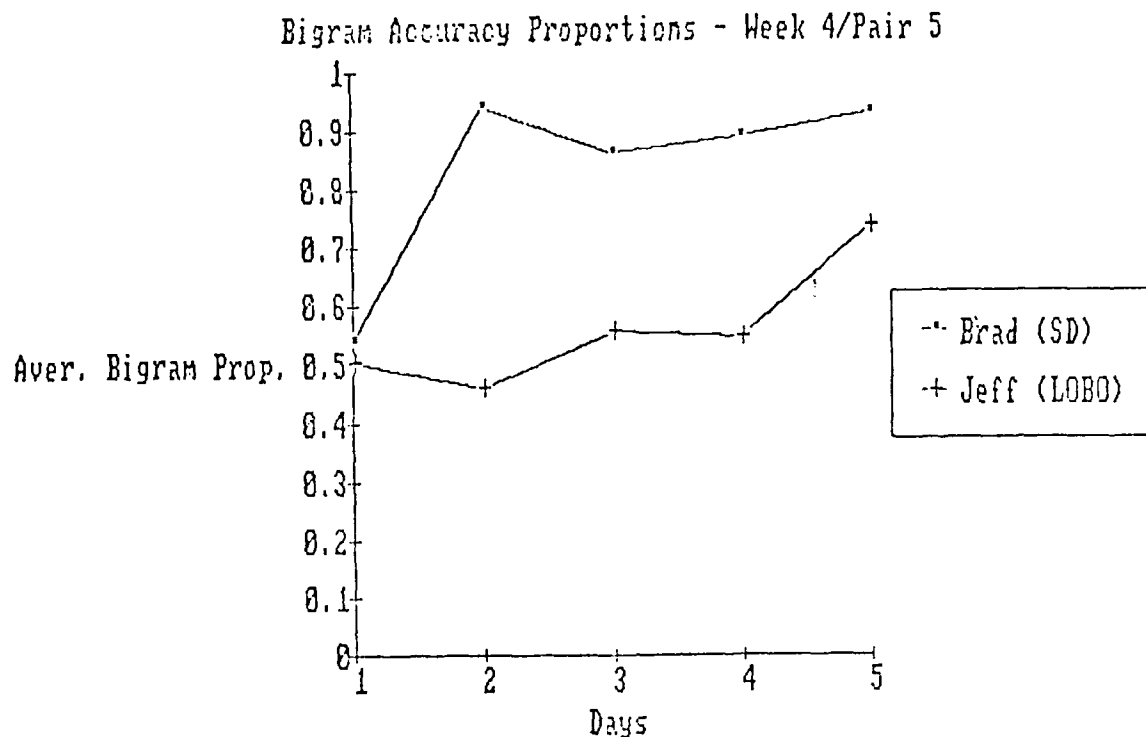
Figure 3. Bigram accuracy proportions for week 5.

Bigram Accuracy Proportions - Week 5/Pair 5
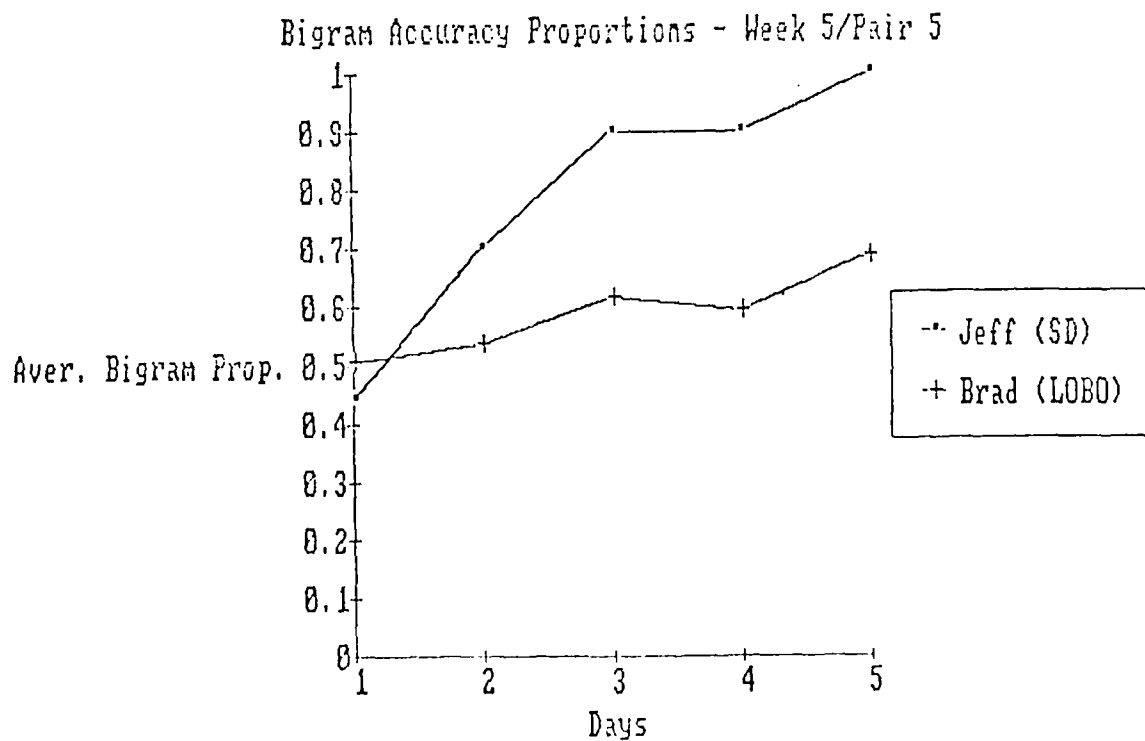
achieved nearly perfect bigram accuracy. Their respective performances using the control version of the program resulted in only slight increases in bigram accuracy by the end of the week, even though each student practiced the words daily.

Discussion

Software developed as part of this study marks an attempt to go beyond what might be accomplished with other, less expensive media. The unique capabilities of computers to store and search vast amounts of data means that specific academic learning histories can be used to inform instructional components in a program. Ultimately, this may result in programs that can be tailored to the aptitudes and motivational levels of individual learners by varying the type and amount of constraint seeking information available at any given point in the problem-solving process. SPELLDOWN represents a first step in that direction.

References

Gerber, M. M. (1985). Spelling as concept-governed problem solving: Learning disabled and normally achieving students. In B. Hutson (Ed.), Advances in reading/language research. (vol. 3) (pp. 39-75). Greenwich, CN: JAI Press.

Gerber, M. M. (1986). Generalization of spelling strategies by LD students as a result of contingent imitation/modeling and mastery criteria. Journal of Learning Disabilities, 19(9), 530-537.

Gerber, M. M., & Hall, R. J. (1987). Information processing approaches to studying spelling deficiences. Journal of Learning Disabilities, 20(1), 34-42.

Hall, R. J. (1980). Cognitive behavior modification and information-processing skills of exceptional children. Exceptional Educational Quarterly, 1(1), 9-15.

Hall, R. J., Gerber, M. M., & Stricker, A. G. (1988). Cognitive Training: Implications for spelling instruction. In J. N. Hughes & R. J. Hall (Eds.), Handbook of cognitive behavioral approaches in the schools. New York: Guilford Publications, Inc.

Hall, R. J., & Stricker, A. G. (1988, January). Intelligent Computer-Assisted Instruction (ICAI) in Spelling: Effectiveness of scaffolded and faded cueing on the performance of learning disabled students. Paper presented at the annual meeting of the Southwest Educational Research Association, San Antonio, TX.

Kolligian, J., & Sternberg, R. J. (1987). Intelligence, information processing, and specific learning disabilities: A triarchic synthesis. Journal of Learning Disabilities, 20, 8-17.

Stricker, A. G., & Hall, R. J. (1988, April). Application of an intelligent CAI tutoring system to spelling instruction for learning disabled students. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.

Swanson, H. L. (1987). Information-processing theory and learning disabilities: An overview. Journal of Learning Disabilities, 20, 3-7.

Wenger, E. (1987). Artificial intelligence and tutoring systems. Los Altos, CA: Morgan Kaufmann.

EQUAL OPPORTUNITY IN THE DEPARTMENT OF DEFENSE:
PAST, PRESENT, AND FUTURE

Ms. Leslie Wilson
DEOMI Research Division
Lt Col Mickey Dansby
Director, DEOMI Research and Evaluation
Dr. Dan Landis
University of Mississippi

The Defense Equal Opportunity Management Insititute (DEMOI) is responsible for training equal opportunity (EO) advisors from all Services and for conducting research on EO issues. We propose a panel session exploring trends in the military EO arena. The format will be a panel with a moderator (Ms. Leslie Wilson, Chief of the DEOMI Research Division). Panelists will include Lt Col Mickey Dansby, DEOMI's Director of Research and Evaluation, and Dr. Dan Landis, Director of the Center for Applied research and Evaluation, University of Mississippi. The panel will also include at least two others from the military EO community.

The DEOMI representatives will present the latest data on EO within the DoD, highlighting trends in representation and utilization of women and minorities. Issues of contemporary concern (e.g., religious accommodation, sexual harassment, and combat exclusion for women) will be discussed. Panelists will discuss future trends and policy changes that may influence EO in the future. Implications of the changing demographic composition of the Armed Forces will be considered. Current research and policy plans will be discussed as a springboard for audience interaction with the panel and moderator. A number of DEOMI publications will be available for the audience

FACTORS AFFECTING THE EARLY CAREER DECISIONS OF THE FIRST WOMEN GRADUATES OF THE USAF ACADEMY

PANEL DISCUSSION

Capt Melissa R. Kallett
Panel Chair, USAF Academy, Colorado
Lt Col Frank R. Wood
Executive Development Office,
National Defense University
Capt Hugh A. Forde
Capt Ann C. Miner
Capt Mark W. Patton
Capt Bridgett R. Price
Capt Arthur E. Roffey
Department of Behavioral Sciences and
Leadership, USAF Academy, Colorado

SUMMARY

The experiences of the first women graduates from the Air Force Academy (Class of 1980) are unique in the history of the military and of the sex-integration movement. In 1986 this group reached a major decision point in their Air Force careers. For most their initial commitments were completed and they were deciding whether to continue their service in the Air Force or to resign and pursue interests in the home or in the civilian sector.

This is a discussion of a qualitative study of the factors affecting the career decisions of this group of women. It is an analysis of indepth telephone interviews of all the graduates who could be located (46 of the 97 total graduates). The interview involved seven open-ended questions asking about factors affecting career decisions and their early career experiences.

In general, the sample we interviewed was 30.4% single, and 69.5% married, with approximately 2% divorced. In regard to career field, 19.6% were assigned to flying duties, and 80.4% were assigned to support duties. More important to this research, which attempts to examine how factors affect career decisions, women who are "already out" of the Air Force, state that they "will be leaving" in the near future, and who report they are "probably leaving" constitute approximately 23.9% of the sample.

This is a discussion of preliminary findings. We are describing only trends evident in the interview comments of the respondents. Our purpose at this stage of the research is to uncover issues relevant to this group as they reach an intermediate conclusion on their experiences.

A content analysis of the factors affecting their career decisions revealed family concerns as the most important followed by job concerns and then by career issues. In addition to a content analysis other sections of the study include; a detailed report on family issues; an analysis in the nature of relationships between these women and their supervisors; a description of the interactions between these women and others they work with; an assessment of personality characteristics and self-perceptives of these women; an examination of the early career experiences of this group; an analysis of what the Air Force could do better to enhance their career commitment, and an analysis of what the Air Force Academy could have done better to prepare the women for their Air Force career.

These sections raise some interesting issues worthy of further investigation. While the data suggests how these issues operate to affect career decisions, more work is needed formulating questions and examining how extensively these processes affect the entire sample.

APPLYING MACROERGONOMICS

PANEL DISCUSSION

Dr. Hal W. Hendrick, Chair
Dr. Ogden Brown
University of Denver
Dr. Najmedin Meshkati
Dr. Andrew Imada
University of Southern California

ABSTRACT

Historically in system development, human factors has been applied in the design of the human-machine interfaces of individual workstations *or* at the micro-ergonomic level. While optimizing components of systems, this microergonimic approach allows for the suboptimal designing of the overall system. Macro-ergonomics is a newly-developed, top-down systems approach which first optimizes organizational-machine interfaces so as to insure a totally ergonomically designed system. A macroergonomic design process and major macroergonomic considerations are described.

Macroergonomics: A System Design Approach for Developing More Effective Military Systems

Hal W. Hendrick, Ph.D
University of Denver

Abstract

Historically in system development, human factors has been applied in the design of the human-machine interfaces of individual workstations or at the *micro-ergonomic* level. While optimizing components of systems, this microergonimic approach allows for the suboptimal designing of the overall *system*. Macro-ergonomics is a newly-developed, top-down systems approach which first optimizes organizational-machine interfaces so as to insure a totally ergonomically designed system. A macroergonomic design process and major macroergonomic considerations are described.

Human factors historically has been concerned with the design of controls, displays and workspace arrangements. In system design, the operations to be required of the system typically have been analyzed to identify the specific functions that comprise them. The human factors specialist often enters the design process at this point and applies his or her knowledge of human-machine interface technology to designing specific jobs, integrating jobs into work groups, and then designing specific human-machine interfaces.

Although applied within a systems context, the above described activities actually are at the individual or, at best, subsystem level. They thus represent applications of human factors engineering at the *microergonomic* level. Although microergonomic applications of human factors frequently result in significant improvements in human-machine effectiveness, these applications have not always resulted in effective total system design. A primary reason for this failure is because microergonomics is not a true *systems* approach; it fails to first consider the overall system dimensions and to begin by systematically designing the *organization-machine* interfaces. Once the system's organizational structure has been designed ergonomically, one then can ergonomically design individual jobs and workstations so as to ensure an overall optimal *system* design.

The importance of utilizing a macroergonomic approach to system design has been greatly heightened by the incorporation of microelectronics and computer technology into our modern military systems. The more these systems become "hard wired" the more essential it becomes to first structure the overall system and then to apply this new technology in a manner that is compatible.

Dimensions of Organizational Structure

The structure of a system's organization may be conceptualized as having three components: *Complexity, formalization,* and *centralization.*

Complexity

Complexity refers to the degree of *differentiation* and *integration* that exist within an organization. Three kinds of differentiation are inherent within a system's organizational structure (Robbins, 1983): *Horizontal differentiation* or the degree of departmentalization and job specialization that is designed into the system; *vertical differentiation* or the number of hierarchical levels; and *spatial dispersion* or the degree to which the system's activities are performed in multiple locations. Increasing any one of these increases the system's or-

ganizational complexity. *Integration* refers to the extent to which structural mechanisms are included in a system's design for facilitating communication, coordination and control across the differentiated elements of the system. As the degree of differentiation increases, the need for integrating mechanisms also increases.

Formalization

From a human factors perspective, formalization may be defined as the degree to which jobs within organizations are standardized and thus allow for little employee discretion. The simpler and/or more repetitive the jobs to be designed into the system, the greater is the utility of formalization.

Centralization

Centralization refers to the degree that formal decision-making is concentrated in a given position, unit or level (usually high in the organization) thus permitting employees (usus 'ly low in the organization) only minimal input into decisions affecting their jobs.

Designing the System's Organizational Structure

The design of a system's organizational structure involves consideration of the characteristics of three major sociotechnical system components: The *technology* to be employed in the system's design, the *personnel subsystem*, and the *external environment* that permeates the organization. Each of these major sociotechnical components has been studied in relation to its effect on the elements of organizational structure, and empirical models that can be used as design guidelines have emerged.

Technology

Perrow: Knowledge-based technology. Of the various empirically developed models of the effect of technology on organizational structure, the one most widely validated and generalizable is that of Perrow, which utilizes a *knowledge-based* concept of technology (Perrow, 1967). Perrow classifies technology by the action one performs upon an object in order to change that object. This action requires some form of technological knowledge; hence, technology can be categorized by the required knowledge base. Using this approach, he identified two underlying dimensions of knowledge-based technology: *Task variability* or the number of exceptions encountered in one's work, and *task analyzability* or the type of search procedures one has available for responding to task exceptions. These search procedures can range from "well defined" where exceptions can be solved using logical and analytical reasoning, to "ill defined" where problem solving must rely on experience, judgment and intuition. The combination of the two dimensions, when dichotomized, yields the matrix shown in Table 1. Each cell represents a different knowledge-based technology.

a. *Routine* technologies have few exceptions and well defined problems. Routine technologies are best accomplished through standardized coordination and control procedures, and are associated with high formalization and centralization.

b. *Nonroutine* technologies have many exceptions and difficult to analyze problems (e.g., aerospace operations). Most critical to these technologies is flexibility. They therefore require decentralization and low formalization.

59

c. *Engineering* technologies have many exceptions, but they can be handled using well defined rational-logical processes. Thus, they lend themselves to centralization, but require the flexibility that is achievable through low formalization.

Table 1. Perrow's Knowledge-Based Technology Classes

Problem Analyzability	Task Variability	
	Routine With Few exceptions	High Variety With Many Exceptions
Well Defined and Analyzable	Routine	Engineering
Ill Defined and Unanalyzable	Craft	Nonroutine

d. *Craft* technologies typically involve relatively few routine tasks and problems rely heavily on experience, judgment and intuition for decision. Problem solving therefore needs to be done directly by those with the expertise. Thus, decentralization and low formalization are required for effective functioning.

Other models of technology. Other models have been developed empirically that also can be useful in system design. Among the better known of these are Woodward's production technology (1965) and Thompson's technological uncertainty (1967).

Personnel Subsystem Characteristics

The two most important personnel subsystem dimensions to consider in organizational design are the *degree of professionalism* and *psychosocial characteristics* of the employees.

Degree of professionalism refers to the formal training and education of the employees. As persons undergo formal training and education or professionalization, the socialization process creates internalized formalization. As this happens, the utility of a highly formalized organizational structure decreases. From a human factors standpoint, in system design there is a tradeoff between formalizing the organizational structure and professionalizing the jobs and related person-machine interfaces.

Psycho-social characteristics of the system's work force also can significantly interact with organizational structure. For example, Hendrick (1979, 1981) has found evidence to suggest that relatively concrete functioning work groups and managers function best under high centralization, vertical differentiation, and formalization. In contrast, abstract or cognitively complex persons seem to function best with low centralization, low vertical differentiation and little formalization.

External Environment Characteristics

Of the three sociotechnical dimensions discussed herein, the one that empirically has been shown to have the strongest interaction with organizational structure is the external environment with which the system interfaces. Of particular importance are the degrees of environmental complexity and change, which together determine the level of *environmental uncertainty.* Classic studies by Burns and Stalker (1961), Emery and Trist (1965), and Lawrence and Lorsh (1969), among others have shown low vertical and

horizontal differentiation, low formalization, and especially, decentralized decision-making (low centralization) to be associated with effective functioning under conditions of high environmental uncertainty; whereas highly differentiated, formalized, and centralized structures are effective under conditions of low environmental uncertainty.

Integrating Micro- with Macroergonomic Design

Through a macroergonomic approach to determining the optimal design of a system's organizational structure, many of the characteristics of the jobs to be designed into the system already are determined. For example, horizontal differentiation decisions proscribe how narrowly or broadly individual jobs must be designed; decisions concerning the degree of centralization will determine the degree of decision-discression to be included in a given job's design; the degree of formalization will dictate the extent to which functions are to be routinized in the job. Each of these job design decisions impacts on the attendant hardware and software design applications of the system's technology--particularly the attendant design of the human-machine interfaces. For example, the degree of formalization and centralization designed into a given position determines the information requirements for that position; this, in turn, drives the design of the information and decision support systems for that work station, and attendant controls, displays and workspace arrangements.

In summary, effective macroergonomic design drives much of the microergonomic design of the system, and thus ensures optimal compatibility of system components with the system's overall structure. The result is a greater assurance of optimal system functioning and effectiveness.

References

Burns, T. & Stalker, G. M. (1961). *The management of innovation*. London: Tavistock.

Emery, F.E., & Trist, E. L. (1965). The causal texture of organizational environments. *Human Relations, 18*, 21-32.

Hendrick, H. W. (1961). Differences in group problem solving as a function of abstractness. *Journal of Applied Psychology, 64*, 518-525.

Hendrick, H. W. (1981). Abstractness, conceptual systems and the functioning of complex organizations. In G. W. England, A. R. Negandhi, & B. Wilbert (Eds.), *The functioning of complex organizations*. Cambridge, MA: Oelgeschiager, Gunn, and Hain. (pp. 25-50).

Lawrence, P.R., & Lorsch, J. W. (1969). *Organization and environment*. Homewood IL: Irwin.

Perrow, C. (1967). A framework for the comparative analysis of organizations. *American Sociological Review, 32*, 194-208.

Robbins, S. R. (1983). *Organizational theory: The structure and design of organizations*. Englewood Cliffs, NJ: Prentice-Hall.

Thompson, J.D. (1967). *Organization in action*. New York: McGraw-Hill.

Woodward, J. (1965). *Industrial organization: Theory and practice*. London: Oxford University Press.

Applying Participatory Ergonomics as a
Macroergonomic QWL Strategy

Ogden Brown, Jr., Ph.D.
University of Denver

Abstract

This study examines a participatory quality of work life pro-
gram in a company which is a major Department of Defense contractor.
It demonstrates that macroergonomics is not only a design consider-
ation, but it is also concerned with the operation and management of
an optimal organizational system.

American managers do not, in general, seem to utilize participative man-
agement approaches. Many familiar with the "language" of participative man-
agement fail to translate this language into managerial behaviors, and organi-
zational commitment to participatory practices thus seems uncommon (Sashkin,
1982). The participatory quality of work life (QWL) program employed by a
major aerospace company (to which I have given the pseudonym "Aspco" to pre-
serve anonymity) clearly does employ participatory ergonomics as a major in-
strument of company policy (Imada, Noro, & Nagamachi, 1986).

Background

A variety of strong external environmental forces are encouraging the or-
ganization to squarely address the issues affecting quality of work life and
organizational effectiveness.

Environmental Forces

High growth markets, international market conditions, and success against
foreign competition are among such issues as are the realities of societal and
technological change and the need to successfully manage them. Further, com-
position of the labor force is vastly different than in the past. Today there
are more working women and older workers. There are the "baby boomers", now
in their thirties, who are better educated, less likely to accept traditional
authority, and who want growth, fulfillment, and a voice in the decisions
which affect them. Add to these forces changes in legal and economic environ-
ments, and the sum would appear to be an increasingly dissatisfied labor force
(Brown, 1985a). The differential between what peor_e want at work and what is
available is increasing.

Alternative Approaches

Traditional approaches to managing organizations simply do not work any
longer. QWL is no longer just a desirable outcome, it is a necessary one.
Even though quality of work life and organizational effectiveness are indepen-
dent outcomes, they are the principal objectives of most organizations. They
can, and should, go hand in hand (Brown, Hendrick, Imada, & Van Slyke, 1985).

A growing number of new beliefs, policies, and practices have come into being, and today there are many approaches to quality of work life.

Parallel Organization

A research project conducted by Goodmeasure, Inc. has identified 21 work alternatives or styles of QWL (Stein, 1983). The alternative utilized by "Aspco" is one grounded in participatory approaches which allow lower level managers and workers more influence in problem solving and governance of the organization. This alternative approach is that of the parallel organization: a management steering committee which guides a set of worker-manager action groups towards solutions of company wide problems (Kanter, 1983). The steering committee is identical to the President and his direct staff. From inception, then, top level support, commitment, and direct involvement were assured (Brown, 1985b).

"Commitment to Excellence" Program

The "Aspco" decision to embark on a major performance improvement effort resulted in establishment of the Commitment to Excellence Program in late 1984. It is quite comprehensive and permeates the entire company. The stated objectives of CTE are to maintain and enhance the competitive position of the company and to provide a work environment which will ensure the quality of work life. The CTE Program is divided into six parts which are interrelated and mutually supportive. The common focus is performance improvement and improved QWL. The program reaches every aspect of "Aspco" business operations (Commitment to Excellence Newsletter, 1985, Nov. 13). A strength of the program is in maintaining a proper balance among the six parts so that no one is either overemphasized or neglected.

Awareness

This portion of CTE is concerned with general and specific actions to raise the level of interest and understanding concerning the importance of performance improvement as a means to realize organizational effectiveness and quality of work life. Steadily increasing quality of performance is the key to realization of both personal and corporate objectives. Every division of "Aspco" has acted to promote awareness of the CTE Program. A periodic Commitment to Excellence Newsletter is published which details significant CTE developments and accomplishments, and informs managers on concepts and techniques for performance improvement. The publication goes to about 1,400 managers, with secondary distribution to employees.

Employee Involvement

Participation by all employees is encouraged through such activities as an employee suggestion program, quality circles, individual training, and recognition awards. An employee suggestion program was begun in the fall of 1985 and resulted in about 1,000 suggestions in the first year, about one-third of which were processed. About twenty per cent were adopted: most of the adopted suggestions earned awards for the originators. Quality circles were begun in 1978 and there are about 75 in existence as of last year. An evaluation in 1985 reflected cumulative savings greater than $10 million by the end of 1984 with a savings to cost ratio of 10:1. An interim study for

1985 showed approved savings of $2.7 million with a potential savings of another $2 million. The effectiveness of properly implemented and managed quality circle programs has been well documented (Brown, 1983, 1985b; Lawler & Mohrman, 1985; Imada, Noro, & Nagamachi, 1986; Lawler, 1986).

Management Training

Managers at all levels receive training to provide a working knowledge of the principal methods of Deming, Juran, Tice and others. Such training includes lectures, seminars, formal in-house courses, on-the-job training, and the like. An arrangement with a major university affords "Aspco" personnel the opportunity to pursue a graduate management degree for which the company pays the entire cost. This amounts to an average annual outlay for "Aspco" of approximately $360,000, an indication of the company's commitment to improving performance and quality of work life. "Aspco" has also emphasized Kepner-Tregoe training in problem solving techniques. Over 2,000 have received this training. In 1986, a seminar series entitled "Investment in Excellence" was begun and expanded in 1987. A new in-house, 25 hour program entitled "CTE Manager's Course" was created and installed last year.

Vendor Involvement

This portion of CTE includes communication to vendors of "Aspco's" emphasis on quality and aid to vendors in the initiation and implementation of more effective quality methods. It is also a positive trend toward improving communications with vendors to reflect "Aspco" interest in their overall performance, and to acknowledge and reward good performance on their part. Almost half the value of "Aspco" is in procurements from vendors and subcontractors.

Customer Outreach

This portion of the CTE Program involves communications with the client to assure understanding of the client's needs, and to generate confidence that "Aspco" is a reliable source of products which meet requirements at a reasonable cost.

CTE Projects

A major objective of the CTE Program is to focus on major chronic problems which may detract from quality and productivity. As part of CTE, a set of problems is identified each year as "CTE Projects," to be subjected to an organized effort to reach feasible solutions as part of the annual program of performance improvement. The annual goal is 12 projects.

It is envisioned that many hundreds of problems will be under study and correction at any given time as "Aspco" institutionalizes the identification and removal of barriers to productivity and quality. Every major division of "Aspco" has developed a set of CTE projects, which they term "problems scheduled for solution," and which they are presently pursuing. Over two hundred CTE projects are under way at present, some of which cross division and other organizational boundaries within "Aspco." As an example, the Production Operations Division has a project with far reaching implications on Computer In-

tegrated Manufacturing, and Strategic Systems has a video-conferencing project which applies to the entire organization.

Certain CTE projects have been identified by the President of "Aspco" as "Top Level CTE Projects." In this category are company-level projects, involving more than one division and requiring a multidisciplinary approach. Each project was a process or activity free of immediate crisis, but which appeared to offer the opportunity for significant improvement. Because of the scope and potential of such projects, leaders are designated from the vice presidential level and are authorized to select team members from any appropriate segment of the company. As an example, a subcontract management project resulted in centralization of this function. Previous thinking was that about 10 to 15 per cent of the company's business was outside business. The CTE project revealed that about 40 per cent of business was with outside vendors which caused "Aspco" to look both ways in contract management. A major reorganization resulted which set up a new staff entity and is now a central function for subcontract management. This reorganization has now offered a central means to career development within this subfield.

CTE Program Administration

The Commitment to Excellence Program is meant to be internalized and implemented within the existing organizational structure as an integral part of daily operations. Each major division is required to plan, organize, direct, and control its portion of the program in keeping with an overall general plan and within broad guidelines from the Central CTE Steering Committee which is chaired by the President of "Aspco" and which includes his 12 vice presidents. This committee establishes policy, approves plans and budgets, and reviews the CTE program on a monthly basis.

CTE Working Groups

CTE Working Groups, within each division, assist the senior manager in defining and implementing the CTE Program for that organization. A great deal of flexibility exists to adjust the program to special needs, but each relevant portion of the six categories of performance and quality of work life improvement is expected to receive attention. Each major division is required to identify a person to assist the manager on CTE matters and to be the division representative to the CTE Council. The Council is composed to designated representatives from each major division of "Aspco." The primary function of the Council is to provide the opportunity for exchange of ideas between major divisions in connection with performance and quality of work life improvement. The Council also serves as a means for communication of information on productivity and quality improvement concepts and techniques.

Analysis of the CTE Program

Considerable progress has been realized in both performance improvement and quality of work life under the CTE Program. It is readily apparent that participatory ergonomics has been a success. "Aspco" has the advantage of being a strong organization with a long history of success in the aerospace industry and a corporate culture of "mission success." The additional value of CTE is that it focuses upon the effectiveness of the processes which bring about the end result and recognizes the need for reaching the objective

directly. By means of the parallel organization approach to the enhancement of productivity and quality of work life (that is, the CIE Steering Committee, the CIE Council, and the CIE Working Groups within each major division), "Aspco" seems to have accomplished a great deal toward realizing these objectives. Keys to this success appear to be: the parallel organization structure; a steering committee of top management overseeing the process diagnosis and problem identification involving the entire company; a "depth strategy" of forming participatory groups to address important issues; and the fact that actions enjoy legitimacy (i.e., adequate authority and empowerment is provided so that CIE actions are perceived as legitimate).

Summary

"Aspco" is committed to a long range program which includes participatory methods and organization-wide change. This program appears to be well managed by those who understand and accept the strategies and values of such approaches. There is a climate of trust, mutual respect, and a firm commitment on the part of "Aspco" from the Chief Executive on down to employ many alternative management strategies, including participatory approaches, to improve product quality, productivity, and quality of work life (Brown, 1986).

References

Brown, O. Jr. (1983). Quality control circles: panacea, program, or philosophy? Proceedings of the Human Factors Society 27th Annual Meeting (p. 1043).

Brown, O. Jr. (1985a). Evolutionary changes in America's work force. In W.A. Hamel (Ed.), Proceedings of the National & Western Regional Conference of the Association of Human Resources Management & Organizational Behavior (pp. 111-116).

Brown, O. Jr. (1985b). Some perspectives on participatory ergonomics in high technology organizations. Proceedings of the Human Factors Society 29th Annual Meeting (pp. 880-883).

Brown, O. Jr. (1986). Participatory ergonomics: historical perspectives, trends, and effectiveness of QWL programs. In O.Brown, Jr. & H.W. Hendrick (Eds.), Human factors in organizational design and management II (pp. 433-437). Amsterdam: N. Holland.

Brown, O. Jr., Hendrick, H.W., Imada, A.S., & Van Slyke, M.D. (1985). Increasing participatory ergonomics and organizational functioning: a field experiment. In I.D. Brown, R. Goldsmith, K. Coombes, & M.A. Sinclair (Eds.), Proceedings of the Ninth Congress of the International Ergonomics Association (pp. 877-888).

Imada, A.S., Noro, K., & Nagamachi, M. (1986). Participatory ergonomics: methods for improving individual and organizational effectiveness. In O. Brown, Jr. & H.W. Hendrick (Eds.), Human factors in organizational design and management II (pp. 403-406). Amsterdam: North Holland.

Kanter, R.M. (1983). The change masters: Innovation and productivity in the American corporation. New York: Simon & Schuster.

Lawler, E.E. & Mohrman, S.A. (1985). Quality circles often the fad. Harvard Business Review 85 (1), (pp. 64-71).

Lawler, E.E. (1986). High-involvement management. San Francisco: Jossey-Bass.

Sashkin, M. (1982). A manager's guide to participative management. New York: AMA.

Stein, B.A. (1983). Quality of work life in action. New York: AMA.

MACROERGONOMICS AND TECHNOLOGY TRANSFER

Najmedin Meshkati, Ph.D.

University of Southern California
Los Angeles, California

Abstract

Technology transfer and its role in economic development is studied. Numerous studies of appropriate technology indirectly and implicitly indicate that micro- and macroergonomic considerations play instrumental roles in the success of the technology transfer project. Major micro- and macroergonomic considerations of technology transfer are discussed and it is concluded that only through a proactive and systematic incorporation of these considerations could one ensure the appropriateness, safety, and effective utilization of the transferred technology.

Technology transfer constitutes one of the major tools in achieving the goal of economic development for developing countries. However, technology transfer, without incorporating proper and needed ergonomic considerations, is doomed to failure (cf., Meshkati and Robertson, 1986). In order to ensure the appropriateness, efficient utilization, and success of transferred technology, major ergonomic considerations, as discussed in this work, should encompass both the micro- and macroergonomic aspects of the analysis.

Economic Development and Technology Transfer

Technology is a key factor in economic development. No developing country can afford to reinvent the wheel when much of the needed advanced technology is already available in more highly developed countries. As formulated by the Denver Research Institute (1983), the crucial questions faced by developing countries are which technology to acquire, how to master it, how to adapt and adjust it to specific sources and the environment of a given country, how to maintain it, and how to build upon it.

Appropriate Technology

Transferring 'appropriate technology' is the optimal solution to the problem. Numerous studies attempt to clarify this concept, its required characteristics and benefits. These studies include, among others, works by Evans and Nogg Adler (1979), United Nations Conference on Trade and Development (UNCTAD, 1972), Development Center of the Organization for Economic Cooperation and Development (OECD, 1973).

Stewart (1979), in a study for the World Bank, argued that appropriate technology consists of technology more in line with developing countries' needs and resources. In a subsequent study, she defined a technology as appropriate only if it reflects the environment in which it is to be utilized.

The quest for appropriate technology is also one of the four basic technological objectives of World Bank operations. An exhaustive report entitled *Technology and Science in World Bank Operations (1982)* argues that "the Bank's techniques for preparation and appraising projects are intended to ensure that the technology used is tailored to the social, cultural, and environmental needs of specific situations."

Micro- and Macroergonomics and Technology Transfer

Ergonomics and Technology Transfer

Ergonomics, also called Human Factors, is concerned with optimizing the productivity, health, safety, and comfort of people, and the efficiency of interactive components of human-machine-environment systems. The payoffs of proactive incorporation of and on-line compliance with human factors considerations in technology transfer projects to developing countries are reflected in economic and non-economic gains. They include improvement of functional effectiveness; higher equipment utilization and maintainability, enhancement of human welfare, and minimal adverse environmental effects due to haphazard and sub-optimal human-machine- environment interactions. A recent study by Kogi and Sen (1987) documents the practical success of incorporating ergonomic considerations in a number of real-life situations in the Third World.

Ergonomics, due to its multidisciplinary nature, is capable of systematically identifying and addressing most of the critical factors determining the appropriateness of an industrial technology for a developing country. Works of Sen (1984), Wisner (1985), Shahnavaz (1985) and Meshkati (1986) attest to the vital role of ergonomics in the successful and efficient selection, transfer, and utilization of imported industrial technology by developing countries.

A Representative Case: Ergonomics and the Bhopal Tragedy

Meshkati and Robertson, 1986 examined the vital role of ergonomic considerations in the success (or failure) of four technology transfer projects. One of the cases studied was that of the Union Carbide pesticide plant in Bhopal, India. The catastrophic accident of December 4, 1984 at this plant resulted in the death of approximately 2500 and the injury of 200,000 individuals.

In reviewing what went wrong in Bhopal, industrial experts have placed human factors problems, i.e. errors caused by carelessness, ignorance, confusion, and fright, high on the list (cf., Shrivastava, 1987). This fact is also reiterated in a report by the International Confederation of Free Trade Unions (ICFTU, 1985), which cited serious problems arising from insufficient training of local operators and their high turnover rate. In this regard the Technical Director of the Chemical Manufacturers Association contended: "You can design the best system, but when you deal with people you can create problems." Moreover, an official of the United Nations Environment program, referring to the issue of traditional agriculture-oriented Indian workers and their reaction to the imported technology stated: "They have not internalized the technological culture" (Engler, 1984). The above example is reminiscent of the aftermath of the Three Mile Island Nuclear Power Plant accident (March 28, 1979). It reminds one of the ironic sayings that *Human Factors (Ergonomics), like a vacant seat in a class, is most noted when it is absent.*

Major Micro- and Macroergonomic Considerations in Technology Transfer

Although the above-cited studies of appropriate technology indirectly and implicitly recognize the immense importance of ergonomic considerations in determining the appropriateness of any transferred technology, they stop short of providing a systematic approach to the proactive incorporation of micro- and macroergonomic-based considerations into the study of appropriateness of the respected technology.

In analyzing the major human factors considerations in technology transfer, the author (Meshkati, 1986a and 1986b) identified four major ares of considerations. The first two are mircroergonomially based and the second two are of macroergonomic nature.

Mircorergonomic Considerations

These set of considerations focus on the application and adherence to the findings and principles developed by what is called the "first" and "second generation" of human-machine technology. The first generation of human factors tends to emphasized the design of specific jobs, work groups, and related human-machine interfaces, including controls, displays, workplace arrangements, and work environments (Hendrick, 1987). Among the areas studied are anthropometry and the study of other human's physical, motor and postural capabilities.

The second generation of human factors emphasizes the cognitive nature of work as reflected in system design. This includes considering the end users' cognitive processes and functioning. Included are cognitive capacity and complexity, psychomotor skills, color association, information processing behavior and decision making, and the concepts of mental map and cognitive compatibility.

The two major microergonomically-based considerations are as follows:

1. Ergonomics of the currently used industry-related infrastructural parameters of developing countries, including special physical and environmental variables, anthropometric considerations, physical and psychological characteristics of the labor force, and idiosyncratic Human-Machine and Human-Workstation relationships. A prototypical example for the above is fitting equipment to different user populations. According to Chapanis (1974), this issue of anthropometric differences is more complex that it appears at first. The problem can not be solved simply by taking a piece of equipment built for one population and scaling it up or down for another population. The difficulty is that body proportions also differ among the different people of the world. For instance when applying American anthropometric design standards to other people of the world, it fits nearly 90% of Germans, but only 80% of Frenchmen, 65% of Indians, 45% of Japanese, 25% of Thais, and 10% of the largest Vietnamese (Kennedy, 1975).

2. Technology Adjustment Operations. *This* refers to adjusting the transferred technology based on the user's country's environmental, social and educational characteristics. It is a broad concept and consists of activities at different levels of project design, implementation, and utilization. It covers both the hardware, software, and maintenance (requirement) adjustments as well as development of specialized maintenance manuals, schedules and procedures. A poor quality translation of a maintenance manual (i.e., not conceptually adjusted), for instance, resulted in performance that was about seven times worse that the performance demonstrated by the operators who used the original manual (Chapanis, 1974).

Macroergonomic Considerations

The third generation of human factors, also called macroergonomics, is a concern with the impacts of technological systems on organizational systems. According to Hendirck (1987), it is entirely possible to do an outstanding job of microergonomically designing system's components, modules, and subsystems and yet fail to reach relevant system effectiveness goals. The microergonomic-related efforts are usually at the individual, team, or, at best, subsystem level. However, the third generation of human factors is focused at the macroergonomic or overall organization-machine system level and concentrates on the development and application of Human-Organization-Technology (HOT) interface technology. The macroergonomic approach builds upon the microergonomic guidelines and findings and continues with an assessment of the implementing organization from top to bottom, using a sociotechnical system approach to organizational and system design. This approach is also consistent with the Total System Design concept which has been recently implemented in design and operation of large-scale human-machine systems.

The sociotechnical system concept considers organizations as open systems composed of two primary subsystems; personnel (i.e., human beings) who are the required actors in the performance of work, and technology, which represents the machinery, tools, techniques, and

methods of task performance. The achievement of a desired outcome requires the joint operation of both subsystems. Specifically, it is proposed that "it is impossible to optimize for overall performance without seeking to optimize jointly the correlative but independent social and technological (sub-) systems" (Davis and Trist, 1972, p.3). The term *Joint Optimization* refers to those conditions in which the requirements of the technology and the physical, social and psychological needs of the human operators are jointly met (Cummings, 1977).

Two primary macroergonomic considerations of technology transfer are as following:

1. The Effects of Cultural Variables on Technology Transfer. Culture, according to anthropologists, is the way of life of people; the sum of their learned behavior patters, attitudes, customs, and material things. However, culture could operationally be defended as the "collective mental programming of peoples' minds" (Hofstede, 1980).

If provisions for cultural and religious compatibility with the various aspects of transferred technology is considered in advance, then the result of the transfer would be positive; otherwise the failure is inevitable (cf., Meshkati and Robertson, 1986). Cultural and religious variables' reaction to transferred technology are complex, interrelated, and hard to single out. However, generally, in technology transfer context, it is reflected through such processes as: attitudes toward work, technology, organization, working habits, working group dynamics, career concepts, achievement and complexity orientation, motivational factors, expectation of rewards, and personal utility and preference systems.

The inherent systematical approach and integrative nature of macroergonomics as a discipline and methodology plus understanding the technical aspects and constraints, along with the organizational and business constraints, led Dray and Monod (1986) to suggest there is an importance of macroergonomics in achieving "synergistic solutions" to the introduction of new technologies within organization.

2. Managerial and Organizational (Methods) Transfer. The important role of the managerial and organizational factors in the success of any technology transfer project is acknowledged by many studies on technology transfer including a policy analysis by the Denver Research Institute (1983) for the Agency for International Development (AID). This work concludes that "without suitable and careful analysis of management factors in any technology transfer transaction, the prospect for success diminishes rapidly." Thus, this suggests a proactive and systematic approach to managerial and organizational transfer to accompany technology transfer. Meshkati and Driver (1986) proposed a systematic method for adjusting and incorporating the (technology) receiver countries' organizational and managerial priorities in the transfer of Western managerial and organizational methods.

Conclusion

Industrial technology transfer constitutes one of the most important tools in achieving the goal of economic development. Only through a proactive and systematic micro- and macroergonomic analysis could one ensure the safe and effective utilization of the transferred technology by the receiving country.

References

Chapanis, A. (1974). National and cultural variables in ergonomics. Ergonomics, 17(2), 153-175.
Cummings, , T.G. (1977). Sociotechnical Systems: An Intervention Strategy. In W. Warner Burke (Ed.) Current Issues and Strategies in Organizational Development, New York: Human Sciences Press.
Davis, L.E. and Trist, E.L. (1972). Improving the quality of work life: Experience of the sociotechnical approach. In Work in America, U.S. Department of Health, Education, and Welfare.
Denver Research Institute (1983). International Transfer and Acquisition of Technology, final report. Prepared for: Agency for International Development (AID), Washington, D.C.
Dray, S.M. and Monod, E. (1986). The new "internationalizm" in Macroergonomics. In O.

Brown and H.W. Hendrick (Eds.), <u>Human Factors in Organizational Design and Management II</u>, Amsterdam: Elsevier Science Publishers B.V.

Engler, R. (1985). Technology out of control. <u>The Nation</u>, April 27, 1985, 488-500.

Evans, D.D. and Nogg Adler, L. (eds.) (1979). <u>Appropriate Technology for Development: A Discussion and Case Histories</u>, Boulder, Colorado: Westview Press.

Hendrick, H.W. (1987). Macroergonomics: A concept whose time has come. <u>Bulletien of the Human Factors Society, 30</u>(2), 1-2.

Hofstede, G. (1980). <u>Culture's Consequences: International differences in work related values</u>, Beverly Hills, California: Sage Publications.

International Confederation of Free Trade Unions (ICFTU). (1985). <u>The Trade Union Report on Bhopal</u>. Brussels, Belgium.

Kennedy, Kenneth W. (1975). International Anthropometric Variability and Its Effects on Aircraft Cockpit Design. In Alphonse Chapanis (Ed.), <u>Ethnic Variables in Human Factors Engineering</u>, Baltimore   Maryland: Johns Hopkins' University Press.

Kogi, K. and Sen, R.N. (1987). Third world ergonomics. In D.J. Oborne (Ed.), <u>International Reviews of Ergonomics</u>, London· Taylor and Francis.

Meshkati, N. (1986a). Major human factors considerations in technology transfer to the industrially developing countries: An Analysis and Proposed Model. In O. Brown and H.W. Hendrick (Eds.), <u>Human Factors in Organizational Design and Management II</u>, Amsterdam: Elsevier Science Publishers B.V.

Meshkati, N. (1986b). Human factors of technology transfer: Resurgent interest in an emerging field. <u>Bulletin of the Human Factors Society, 29</u>(12), 1-2.

Meshkati, N. and Driver M.J. (1986). A systematic method to analyze the effects of the non-technological infrastructure factors of the industrially developing countries on managerial effectiveness and success. <u>Proceedings of the 30th Annual Meeting of the Human Factors Society</u>. Dayton, Ohio, September 29 - October 3, 1986.

Meshkati, N. and Robertson, M.M. (1986). The effects of human factors on the success of technology transfer projects to industrially developing countries: A review of *representative case studies. In O. Brown and H.W. Hendrick (Eds.)*, <u>Human Factors in Organizational Design and Management II</u>, Amsterdam: Elsevier Science Publishers B.V.

Organization for Economic Cooperation and Development (OECD), (1973). <u>Transfer of Technology for Small Industries</u>. Proceedings of Seminar Held at Schloss Hernstein, Austria, 5-8 July, 1973.

Sen, R.N. (1984). Applications of ergonomics to industrially developing countries. <u>Ergonomics, 27</u>(10), 1021-1032.

Shahnavaz, H. (1985). <u>Ergonomics of Industrially Developing Countries</u>. Center for Ergonomics of Developing Countries (CEDC). Department of Human Work Sciences, Lulea University, Sweden.

Shrivastava, P. (1987). <u>Bhopal: Anatomy of a Crisis</u>. Cambridge, Massachusetts: Ballinger Publishing Company, A Subsidiary of Haper & Row, Publishers, Inc.

Stewart, F. (1979). International technology transfer: Issues and policy options. <u>World Bank Staff Working Papers</u>, No. 344, The International Bank of Reconstruction and Development / The World Bank, Washington, D.C.

The World Bank (1982). <u>Technology and Science in World Bank Operation</u>. The International Bank of Reconstruction and Development / The World Bank, Washington, D.C.

United Nations Conference on Trade And Development (UNCTAD), (1972). <u>Guidelines for the Study of the Transfer of Technology to Developing Countries</u>, A study by the UNCTAD Secretariat, United Nations, New York.

Wisner, A. (1985). Ergonomics in industrially developing countries. <u>Ergonomics, 28</u>(8), 1213-1224.

PANEL DISCUSSION

**TRAINING RESEARCH & DEVELOPMENT (R&D)
IN THE AIR FORCE**

Colonel Gene A. Berry, Chair
Major Daniel L. Collins
Captain Robert E. Duncan
Captain Amy E. Potts
Lieutenant Charles G. Capps
Air Force Human Resources Laboratory

PURPOSE

To discuss current R & D programs of the Air Force Human Resources
Laboratory which focus on bringing new technologies to bear on current
and emerging Air Force training problems and opportunities.

Intelligent Tutoring Systems

Lt Charles G. Capps
Training Systems Division
Air Force Human Resources Laboratory

Increasingly complex weapon systems require more detailed, dynamic training environments. The Air Force Systems Command (AFSC) has identified Artificial Intelligence (AI) as one of the most promising and pervasive technologies for the future. The Air Force Human Resources Laboratory (AFHRL) has the charter of finding out whether or not AI has a place within the Air Force's technical training environment. The delivery vehicles for bringing AI into the Air Force classroom are Intelligent Tutoring Systems (ITS).

ITS capitalize on the design, implementation, and evaluation of computer-based training (CBT) systems over the last two decades. ITS differs from tradi- tional CBT in three ways. First, the system must be smart enough (contain enough domain knowledge) to draw inferences or solve problems in the domain. Second, the system must be able to encode what the student knows as well as what the student doesn't know. Finally, an ITS must contain the proper instructional pedagogies to tutor the needs of the individual student who is working with the system. ITS are dynamic, real-time learning environments which respond to students as separate learning entities.
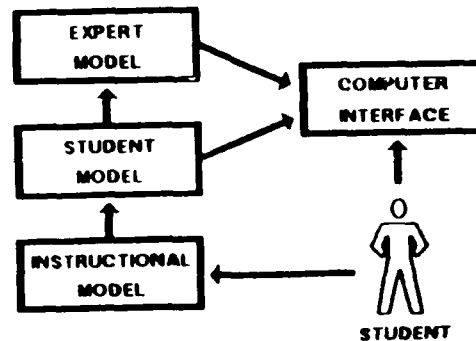
Recent experience has shown that ITS can fill a great need by providing excellent training opportunities for students involved in courses which contain a vast amount of information. The ITS with its expert system, student conceptions and teaching theories gives the student an interactive environment in which he/she can learn and implement, the information presented either by the ITS or by a classroom instructor.

The space environment is an excellent candidate for the implementation of these systems. Two ITS currently ready for reliability and validity studies are the Orbital Mechanics Tutor, designed to aid students at the Undergraduate Space Training School, Lowry AFB CO and the Fuel Cell Tutor for the Space Shuttle at NASA's Johnson Space Center, TX. Empirical evaluations of these systems will be conducted to assess overall cost and training effectiveness and acceptance by students and instructors. ITS promises to be an integral training paradigm for the Air Force's future.

Job Performance Measurement in Training Evaluation

Captain Robert E. Duncan, Ph.D.
Training Systems Division
Air Force Human Resources Laboratory

Classically, training evaluation has focused on the knowledge and skills trained, with only an inferrence that training or changes in training enhance job performance. With few exceptions, determining the effect training interventions have on actual job performance suffers from "the criterion problem" (no adequate measure of performance). In an attempt to rectify this problem and to respond to an Air Training Command (ATC) and Tactical Air Command (TAC) joint need, the Air Force Human Resources Laboratory is attempting to develop measures to evaluate the effectiveness of additional or modified technical training for TAC-bound F-100 jet engine mechanics. This includes measures to determine differences in knowledge and skills gained by additional training (internal evaluation), and measures to determine differences in job performance attributable to the additional training (external evaluation).

The Air Force's Job Performance Measurement System (JPMS), includes several methods for tapping job performance. The Walk-Through Performance Test (WTPT) is comprised of a "hands-on portion", where the incumbent actually performs the task, and an interview portion, where he explains steps required. Rating forms completed by supervisors peers, and incumbent are also used, and job experience is collected to use as a moderator in performance prediction equations.

A joint ATC/TAC initiative will change the jet engine mechanic technical training program by adding to and modifying the objectives of the course to incorporate greater weapon system-specific training. The intent is to increase technical competency of specific graduates (TAC-bound, F100 engine mechanics). The revised training is as an attempt to go from the 3-skill level to a hypothetical 4-skill level. An obvious requirement is to assess the impact of the additional training on knowledges, skills and subsequent job performance.
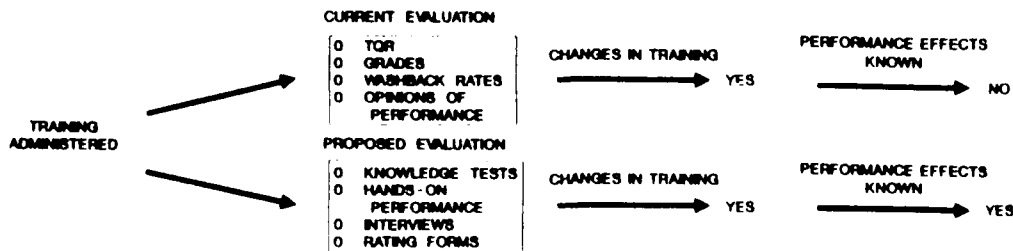
Based on a review of the current jet engine course, 4-level objectives, and tasks performed by F100 jet engine mechanics in their first 12 months on the job, a list of tasks was developed which matched course objectives to tasks performed by newly trained mechanics. From this task list and a list of knowledge-based objectives, written knowledge test items and hands-on and interview performance items were constructed. Performance items followed WTPT construction procedures and were reviewed by ATC and TAC subject-matter experts. As students begin to flow, to the job, trainee job performance will be assessed periodically. Results will reviewed to determine the effectiveness of the 4-level initiative. The use of the job performance measurement technology developed by the Air Force may be a first step in accurately assessing effects of training intervention on job performance.
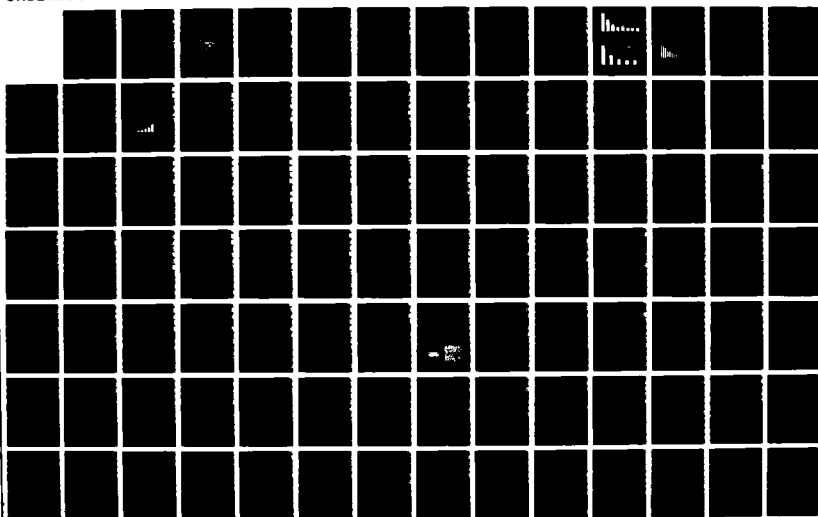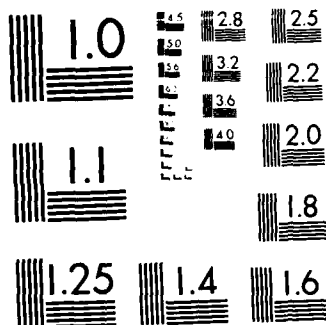
Testing the Effectiveness of a Major Training Innovation

Captain Amy E. Potts
Training Systems Division
Air Force Human Resources Laboratory

A successful on-the-job training (OJT) program is critical to the Air Force's combat mission. As weapon systems have become more technologically sophisticated, the existing OJT program has become expensive to administer and difficult to manage. The Advanced On-the-job Training System (AOTS), is a four-year, three-phased project to build a prototype state-of-the-art training system using the latest computer and automation technologies. The Air Staff directed that the AOTS prototype be developed and evaluated within the operational Air Force environment to insure that the system is both useful and useable for those who would ultimately use it, the NCO force. The prototype is being developed at Bergstrom AFB TX. It will integrate and manage OJT; train and evaluate the individual; and automate job site training to make OJT more responsive to mission requirements. Senior NCOs in four career areas (personnel, security police, aircraft maintenance, and jet engine mechanic) have identified work center training requirements and are now developing training and assessment materials for the system's test and evaluation (T&E).

The 12-month AOTS T&E will begin August 1988. It will involve a year-long, side-by-side comparison of Air Force, Air Force Reserve and Air National Guard work centers, half of each using the AOTS and the other half the current OJT system. The test will include work centers for each of the four AFSs used in the prototype. The evaluation will include both training and operational data as well as the opinions of all personnel involved. The evaluation will involve the CONUS MAJCOMs, Air Force Reserve and Air National Guard, and will be monitored by the Air Force Operational Test and Evaluation Center.
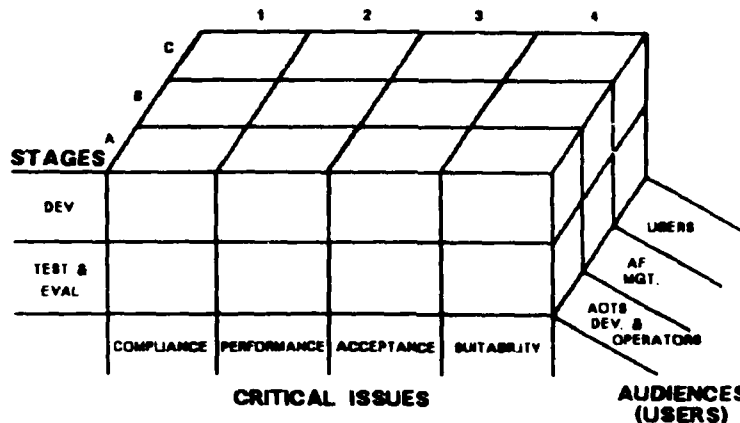
This presentation addresses the test plans and associated problems of accomplishing a major test program in the operational environment. During AOTS development, T&E will be primarily concerned with verifying attainment of technical performance specifications and objectives. This includes testing and evaluating AOTS components, interfaces, and subsystems. Evaluation results can then be used to revise or modify system elements so they peform at or above acceptable levels, and so the integrated system will be functionally complete. T&E will also be accomplished during the operational implementation of the system. Four critical issues will be addressed: compliance, performance, suitability, and user acceptance. The above three dimensional Criteria Acquistion Model will be utilized for the AOTS evaluation.

Decision Support System for Training Management

Major Daniel L. Collins
Training Systems Division
Air Force Human Resources Laboratory

The Training Decisions System (TDS) is being developed to provide a more unified and integrated approach to training planning and management. Specifically, TDS uses information on job tasks performed by airmen, combined with airmen assignment information and Air Force training capacities, to determine cost-effective training options. Recent Air Force budgetary constraints have resulted in a reduced supply of money and personnel for accomplishing Air Force training. Consequently, training decisions in the Air Force are becoming increasingly critical. Furthermore, making those decisions has been made more complex by incomplete and inadequate cost data. Due to the scope and complexity of Air Force training, the challenge has been in deciding what to train (training content), where to train (appropriate training settings e.g., schoolhouse, OJT, field training detachments), and when training should occur (at what point in an airman's career). The resulting decisions, although the best possible given the information available at the time, could have benefited from a large, accurate training data base used in conjunction with optimization models to help answer these questions.
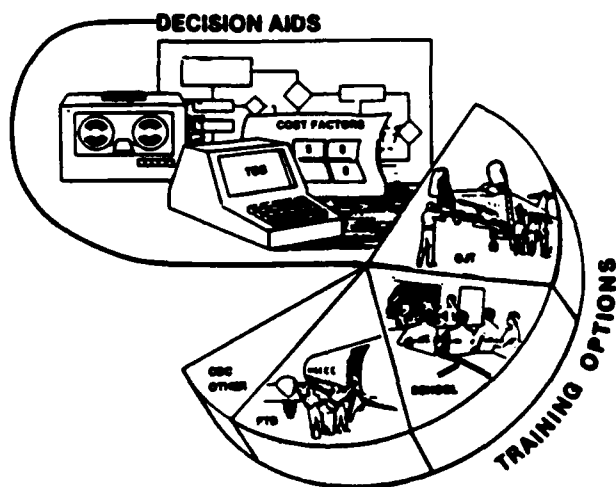
TDS is an extensive multi-year research and development effort consisting of three basic subsystems and a fourth integrating subsystem. The first or Task Characteristics Subsystem, will provide task training modules (TTMs) and associated training site allocation preference data. These TTMs will be the prime building blocks for the other subsystems. The second subsystem, the Field Utilization Subsystem, will provide present and alternative training/personnel assignment patterns for each job specialty and associated preference values. The third subsystem, the Resource/Cost Subsystem, will provide costs and capacity functions which will be an important input to the training optimization routines of the Integration and Optimization Subsystem. This subsystem will result in the integration of the three previously described subsystems. In addition, it will provide optimization software and an interactive system that will allow planners and managers to answer "what if" questions relative to the what, where, and when of Air Force training.

PERSONNEL SECURITY RESEARCH IN DOD

PANEL DISCUSSION

Dr. Martin F. Wiskoff, Chair
Dr. Kent Crawford
Defense Personnel Security Research and
Education Center
Lt Colonel Russell J. Hibler
Major Ronald K. Chapman
Captain Alan R. Kolski
Department of Defense Medical Center,
Ft. Meade, MD
LCDR Forrest Sherman
Marine Security Guard Battalion

ABSTRACT

Personnel security issues have become more prominent in recent years as the press has described in detail the escapades of individuals involved in espionage. As early as 1982 a Defense Panel review of the DoD Personnel Security Program recommended that a research program be initiated with particular emphasis on the potential use of psychological tests and interviews as a supplement to the investigative process. This paper will provide background to the area of personnel security and the evolution of a research capability with DoD. Papers presented on this panel will explore some current DoD research to address personnel security screening and evaluation concerns.

Personnel Security Research in DoD: Program Examples

Martin F. Wiskoff, PhD.
Defense Personnel Security Research and Education Center
Monterey, CA

Abstract

Personnel security issues have become more prominent in recent years as the press has described in detail the escapades of individuals involved in espionage. As early as 1982 a Defense Panel review of the DoD Personnel Security Program recommended that a research program be initiated with particular emphasis on the potential use of psychological tests and interviews as a supplement to the investigative process. This paper will provide background to the area of personnel security and the evolution of a research capability within DoD. Papers presented on this panel will explore some current DoD research to address personnel security screening and evaluation concerns.

In August 1984 Secretary of Defense Weinberger expressed concern with respect to the high number of people cleared by DoD for access to classified information. A comprehensive assessment was undertaken which culminated in an across-the-board 10% reduction in clearances as well as a similar reduction in requests for investigations. Subsequently DoD has achieved a 39% clearance reduction and is seeking further ways to reduce the still sizable number of clearances. But problems still remain.

The Senate Permanent Subcommittee on Investigations in 1985 conducted hearings on personnel procedures to achieve security of government operations. They were dismayed at many of the procedures and expressed concern at the fact that security clearances had been handed out to 4.3 million people. Georgia Senator Sam Nunn viewed this as an unmanageable amount that could not be effectively monitored, and he commented that "the scope of the problem is just unbelievable."

The press has had a field day with stories of espionage in government. In a June 1985 feature article, Newsweek stated that "It has been a dismaying time for the defense establishment. In the last 12 months alone, espionage charges have been brought in eight separate cases, implicating 15 people - including for the first time, an active agent of the FBI..." Two years later, U. S. News and World Report's lead paragraph reads, "Whatever its secret successes have been, the past three years have clearly been the worst of time for America's intelligence community (June 1, 1987). Since January, 1984, no fewer than 20 U. S. citizens have been convicted of stealing national secrets, compared with just four in the previous four-year period."

Most recent publicity has revolved around the Marine Corps and their embassy security guard responsibilities. The nation was stunned by disclosures that the security of the U. S. embassy in Moscow had been violated. While the extent of the infiltration in Moscow and other Eastern bloc embassies has still not been completely determined, the potential implications of these types of security breaches causes great fear and discomfort with our security procedures.

Obviously there cannot be any documentation of the number of undiscovered espionage incidents. Our personnel security investigative system is not fundamentally designed to uncover spies. And we are still not satisfied that it is sufficiently sophisticated to identify the character traits and the kind of misconduct making persons vulnerable to espionage approaches.

As early as April 1982, a Defense Panel review of the DoD Personnel Security Program recommended that a research program be initiated with particular emphasis on the potential use of psychological tests and interviews as a supplement to the investigative process. Three years later the Commission to Review DoD Security Policies and Practices (Stilwell Commission) echoed its support. The report, "Keeping the Nation's Secrets" made the following observations:

Although billions of dollars are spent annually for security, relatively little goes to research activities. Moreover, significant aspects of policy and practice should properly be based upon research. Yet, such research is neither ongoing nor planned.

There logically should be research to determine the optimum structure of background investigations.

There should also be an analysis of the efficacy of the information elicited on personal history statements required to be filled out by clearance applicants; and there should be a similar analytic basis underpinning questions being asked of the subjects by DIS (Defense Investigative Service) investigators. None of this exists.

Adjudication policies also beg for a firmer basis in research.

In February 1986, the Defense Personnel Security Research and Education Center (PERSEREC) was established by the Secretary of Defense to address these and other concerns. PERSEREC's mission is to:

1. Evaluate and improve DoD personnel security procedures, programs, and policies through research and analyses.

2. Provide training, instruction, and advice on selected personnel security subjects to DoD Components.

3. Stimulate joint personnel security research projects that have Defense-wide implications.

4. Identify and collect relevant statistical records and develop data bases for personnel security research.

The PERSEREC charter encompasses research on all DoD personnel, which includes active duty and reserve military, civil service employees and contractor personnel with security clearances. There are four substantive personnel security research areas:

1. Prescreening - The actions taken by agencies to identify and eliminate high security risk individuals prior to the formal processing for a security clearance.

2. Background investigation - The investigation that is conducted as a prerequisite to granting a top secret and above clearance for assignment to highly sensitive positions. The investigation consists of various record checks, i.e. of federal agencies, police, credit, educational and employment sources plus interviews of references and often an interview with the person being investigated.

3. Adjudication - The process by which a decision is made whether to grant a security clearance. The process uses information from the background investigation and possible additional sources and is performed by the agency making the request for clearance.

4. Continuing evaluation - The requirement to reevaluate individuals periodically, to determine whether they should continue to be granted a clearance and at what level.

The issue of personnel security raises a fundamental dilemma in a free society: the need for control of national security versus the equally important requirement to protect the rights of citizens. To some extent these two goals may be viewed as resting on opposite ends of a see-saw, with the weight of governmental and public pressure at any time tipping the balance one way or another.

Panel members will address three personnel security issues today. Dr. Hibler will discuss pathological indices in officers who hold high security clearances and the implications of such a situation. Dr. Sherman will address research and operational procedures, initiated as a result of the espionage incidents in Moscow, to better screen and evaluate Marine Security Guards. Dr. Crawford will present the results of a study to evaluate biographical data as an enlisted personnel security prescreening measure. The panel will discuss the current impetus for personnel security research within DoD and the prospects for the future.

References

A family of spies. (1985, June). Newsweek. pp. 32-38

Department of Defense, Select Panel Review of the Department of Defense Personnel Security Program, April, 1982.

Department of Defense Directive 5210.79, Defense Personnel Security Research and Education Center (PERSEREC), February 19, 1986

Department of Defense Security Review Commission. (1985). A report to the Secretary of Defense by the Commission to review DoD security policies and practices. Washington, DC: Office of the Secretary of Defense.

To catch a spy. (1987, June). U.S. News, pp. 20-28.

U. S. Senate, Federal Government Security Clearance Program. Hearings before Permanent Subcommittee on Investigations, Ninety-Ninth Congress, April, 1985.

Pathological Indices in Personnel With High Security Clearances[1]

Russell J. Hibler
Alan R. Kolski
Ronald K. Chapman
United States Air Force

Abstract

Pathological behavioral indices of 122 military personnel assigned to high level security positions in a large tri-service organization were examined. These individuals were referred for behavioral-security evaluations following episodes of misconduct, or as a result of information uncovered during pre-security access screening. Normative factors are summarized in addition to demographic parameters, applicable DCID 1/14 criteria for security adjudication, screening methods used, DSM-III diagnoses, and recommendations. Implications for personnel security assessment, management, and research are discussed.

The costs resulting from misconduct by individuals in highly sensitive positions, especially through espionage, have been extreme. In most cases active duty, retired military, or civilian employees have successfully met the requirements for security access, but later turned traitor. The base rate for espionage is approximately one for every 300,000 individuals with a security clearance, which makes the prediction of espionage a formidable task (U. S. Senate, 1986; House of Representatives, 1987). Instead, the selection of personnel who are best suited for high responsibility duties is most often accomplished by eliminating those who are considered unsuitable for such trust. This is done primarily through conducting background investigations and interviews. While these methods are effective in screening out individuals with obvious behavior or judgment problems, individuals are often involved in unsuitable behavior after being cleared for security access. For instance, during a recent five-year period 27,000 enlistees, or about 20% of those receiving background investigations for sensitive positions, were later discharged due to subsequent unsuitable behavior or other problems previously undetected (Flyer, 1985). Within the security system, the Director of Central Intelligence Directive (DCID) 1/14 (1984) provides 11 guidelines, or behavioral standards, by which individuals with Sensitive Compartmented Information (SCI), or the highest level of, security access are adjudicated. Despite these findings and procedures, there is little empirical data concerning those who violate the access criteria, the use of psychological screening and assessment methods, or the disposition of these cases. This study examines these factors in individuals who have been in, or considered for, positions of SCI access, and who have exhibited inappropriate behavior.

---

[1] The ideas in this paper are those of the authors and do not necessarily reflect the official policy of the Air Force or the Department of Defense.

Method

Subjects

The subjects were 122 active-duty, tri-service military members who held, were in suspension of, or were being considered for SCI security access. They were referred for evaluation between March, 1983 and December, 1987, and had not been previously evaluated for entry into existing alcohol/drug rehabilitation programs. Subjects ranged in age from 19 to 43, included 104 males and 18 females, and ranged in grade from E-2 through O-3. All had over one year of technical training. The average time in service was 7 years and 2 months, and the average time on station was 1 year and 7 months.

Procedure

Demographic data were obtained through self-report. From two to six one-hour sessions were conducted with each subject covering developmental, family, employment, and military histories, recent life events, mental status examination, and psychological testing. Psychological measures included the Minnesota Multiphasic Personality Inventory (MMPI), the Millon Clinical Multiaxial Inventory (MCMI), the Fundamental Interpersonal Relations Orientation-Behavior (FIRO-B), and the Shipley Institute of Living Scale. The subject's chain of command, security office, medical records, alcohol rehabilitation records, coworkers, family members, and medical specialties were consulted as required. The findings developed by these methods were then used to assess the individual's behavior in relation to the 11 DCID 1/14 behavioral areas. Applicable DSM-III (1980) diagnoses were also provided. Finally, disposition recommendations were made ranging from psychologically clearing the subject for security access to recommending that access be denied.

Results

Sixty of the subjects, or 49%, were referred for evaluation within their first year on station (Figure 1). A comparison of males and females referred for evaluation to the total number of males and females seen in the clinic, including self-referrals, was significant, suggesting a higher evaluation referral rate for males, $X^2=9.92$, df=1, p<.01. There were no sex differences in the rate of occurrence for the five most frequent DCID 1/14 areas (Figure 2). In each of the four age groups, emotional disorders was the most often identified DCID 1/14 area (Figure 3). The screening methods which contributed to a recommendation that access be denied are shown in Figure 4. A comparison of individuals with elevated MMPI clinical scales (>70T) to those with non-elevated MMPI clinical scales, suggested that elevated MMPIs do not differentiate those receiving a recommendation to deny access from those retaining access, while non-elevated MMPIs were more likely to result in a recommendation that access be retained, $X^2=7.4$, df=1, p<.01. Sixty-one of the subjects received a DSM-III diagnosis on Axis I or II: 46 received Axis I, 8 received Axis II, and 7 received both Axis I and Axis II diagnoses. Substance abuse disorders and adjustment disorders were the most frequent Axis I diagnoses, while passive-aggressive personality disorder was the most frequent Axis II diagnosis. Seventy-five subjects were favorably recommended for retention, or awarding, of SCI access.

Months on station
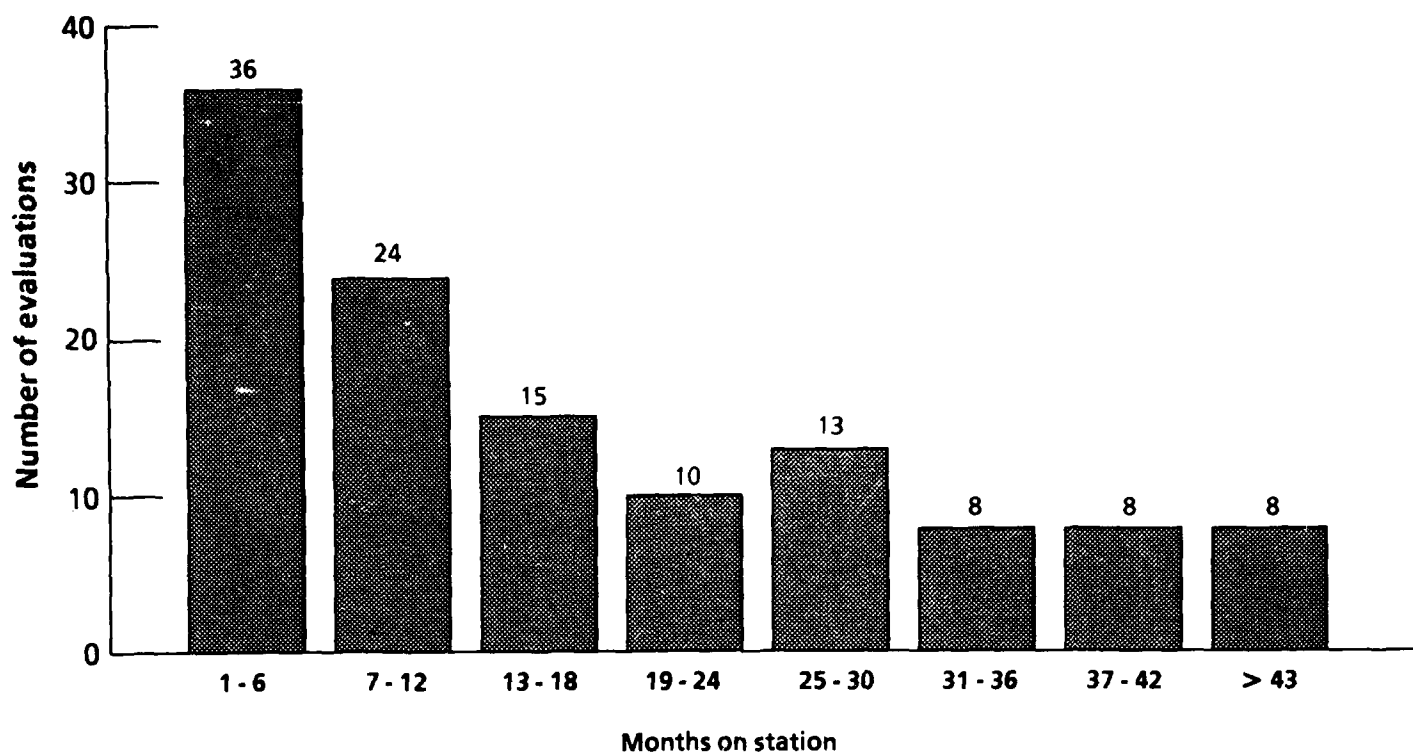
Figure 1. Evaluations as a function of time on station
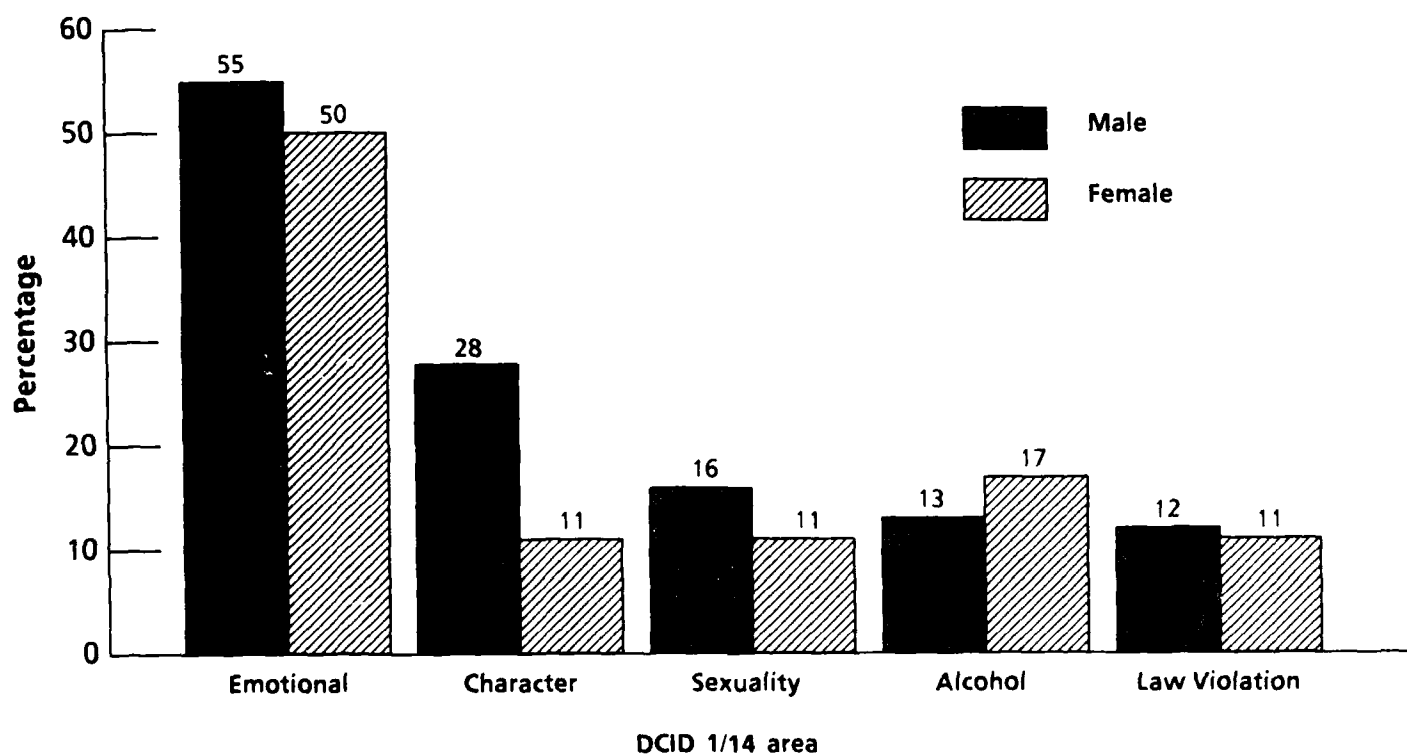
DCID 1/14 area

Figure 2. Five most frequent DCID 1/14 adjudication areas by sex
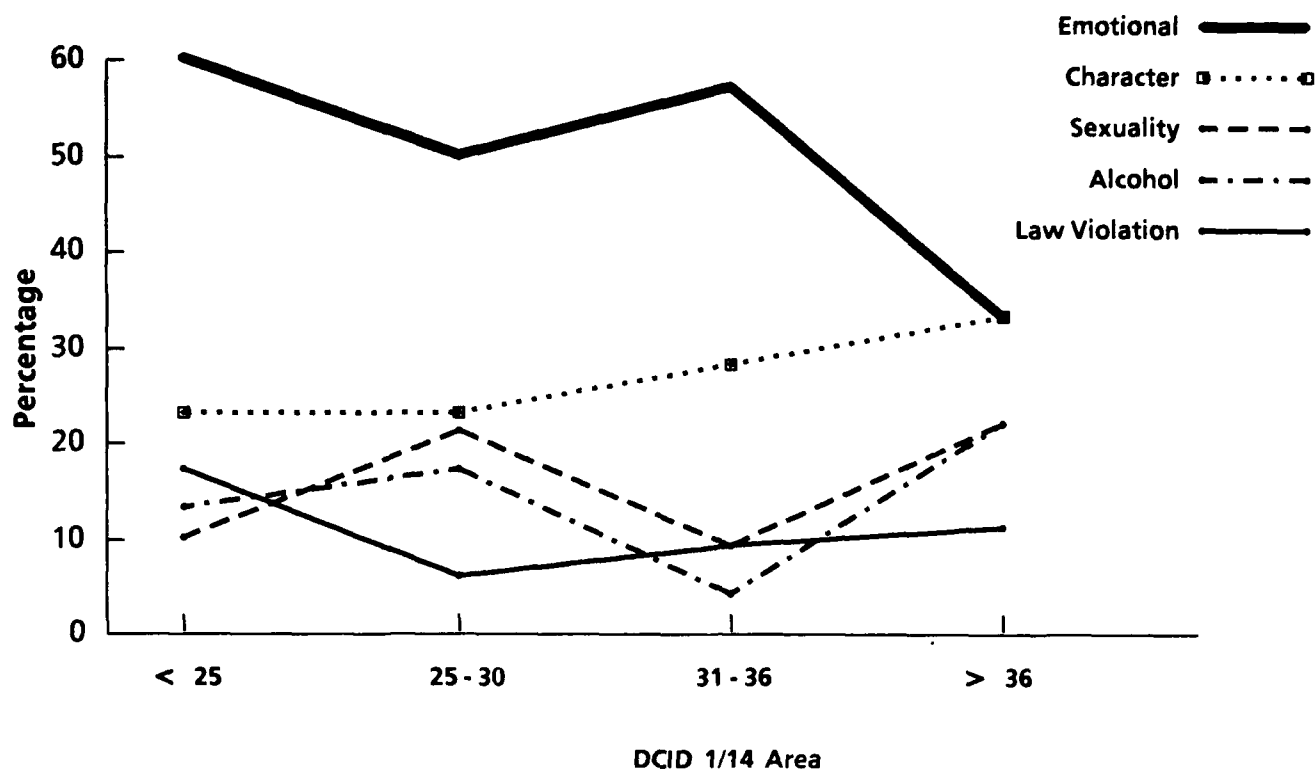
Figure 3. Five most frequent DCID 1/14 adjudication areas by age group
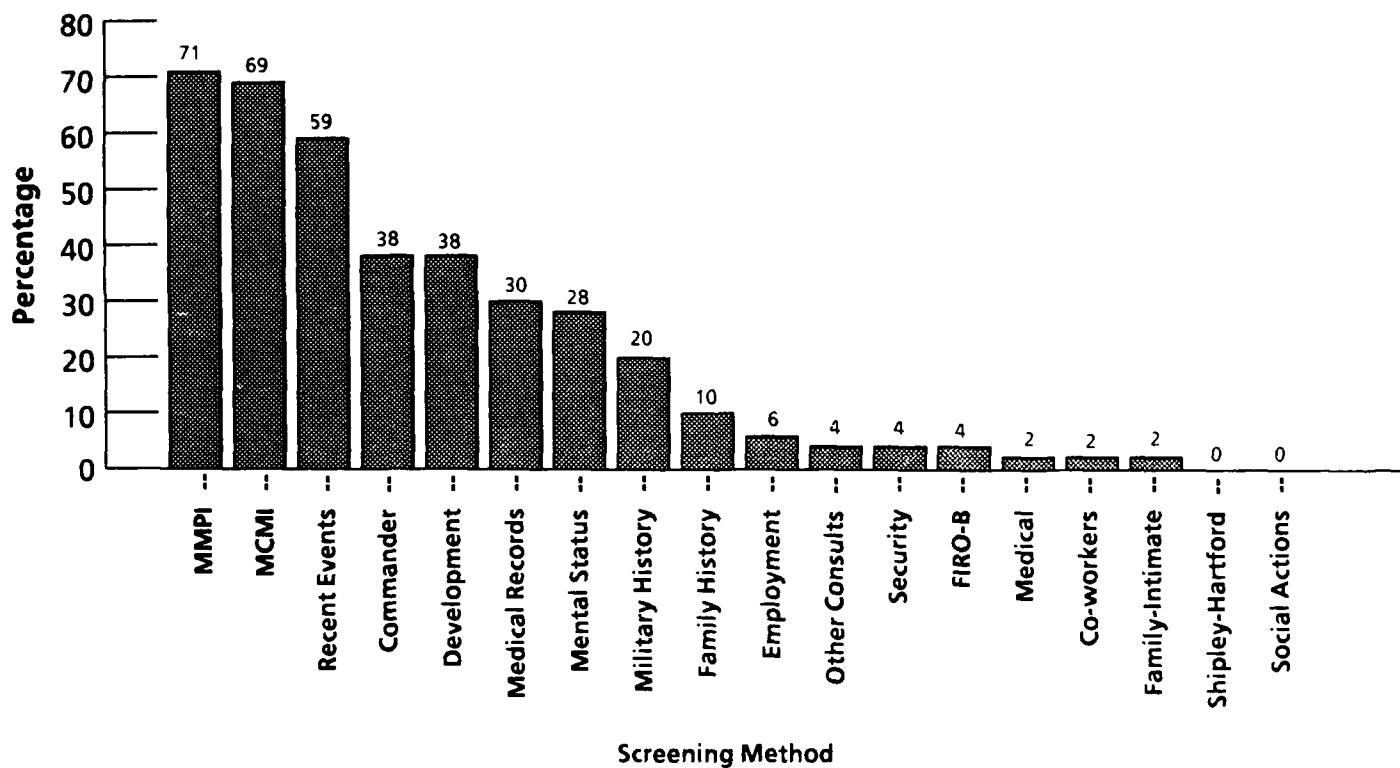
Figure 4. Individuals not recommended for security access by screening method

84

Discussion

Both personality and situational factors affect the decision to deny or recommend the awarding of security access. For instance, it appears that newly assigned individuals are at much higher risk for being involved in security-relevant misconduct than are individuals later in their tours. Almost half of the referrals occurred during the individual's first year, establishing a critical period of vulnerability. This suggests that perhaps newcomers should receive the majority of our leaders' attention and support programs. Within the DCID 1/14 behavioral areas, the emotional and mental disorders predominate for subjects of both sexes. Many of the problems in the emotional area are due to life adjustment and interpersonal issues which might be prevented through training in communication skills, crisis intervention, and a supportive environment. In addition, despite well-established alcohol identification and treatment programs within the military, this study still identified alcohol abuse as a primary concern, one that had gone undetected by existing monitoring systems. More than half of the subjects in this study were recommended for a return to full security access, suggesting a potential cost benefit to be gained from the use of behavioral science consultants to assess the severity of an individual's behavior as it relates to the issue of national security. The use of several screening methods contributed to identifying individuals with potential problem areas and in focusing the clinical inquiry. The objective personality tests, in particular, were most useful for identifying significant pathology. The present study demonstrated the utility of a variety of psychological techniques to assess individuals for security access. By considering many factors it becomes possible to identify relevant situations, timing, and traits, and to identify individuals on whom to focus management resources. Further research on large groups of individuals in a variety of locations is needed in addition to follow-up on the continued performance of those individuals who were evaluated and recommended for security access.

References

American Psychiatric Association. (1980). Diagnostic and statistical manual of mental disorders (3rd ed.). Washington, DC: Author.

Director of Central Intelligence Directive (DCID) No. 1/14. (1984). Minimum personnel security standards and procedures governing eligibility for access to sensitive compartmented information. Washington, DC: Author.

Flyer, E. S. (1985, April). Statement provided before U. S. Senate permanent subcommittee on investigations.

House of Representatives. (1987). Report by the permanent select committee on intelligence (Report No. 100 S). Washington, DC: U. S. Government Printing Office.

U. S. Senate. (1986). Report of the select committee on intelligence (Report No. 99-522). Washington, DC: U. S. Government Printing Office.

BIODATA AS A PERSONNEL SECURITY PRESCREENING MEASURE

Kent S. Crawford

Defense Personnel Security Research and Education Center

Problem and Background

The military services have sensitive billets for enlisted accessions that must be filled only by extremely trustworthy and reliable personnel. These jobs require access to top secret(TS) and/or sensitive compartmented information(SCI). To ensure that newly assigned personnel are qualified for these billets from a security perspective, each service uses various personnel security prescreening procedures. In addition, for candidates successfully passing the initial prescreen, an intensive and costly background investigation(BI) is conducted by the Defense Investigative Service and adjudicated by the respective service.

There is evidence that the above screening procedures are not extremely effective in terms of reducing unsuitability discharge rates. That is, a large number of enlistees successfully passing the service prescreen as well as the BI adjudication are later discharged from sensitive positions for unsuitability. Flyer (in press) found that during the 10-year period from 1976 to 1986, over 50,000 enlisted personnel from all the services who received BIs for TS and SCI clearances were later discharged for failure to meet minimum behavioral standards.

Flyer(in press) also examined unsuitability attrition rates for recruits enlisting during the period 1978 to 1982. Summing across all four services, he found that the total unsuitability attrition rate for personnel entering intelligence specialties was 19.4 percent and the rate for personnel entering other highly sensitive positions was 23.2 percent. These rates were only slightly lower than the 27.6 attrition rate for the total accession population.

While any first-term attrition has costs associated with it, unsuitability attrition from sensitive positions also has security implications. Many of these individuals possess knowledge of highly classified equipment and procedures while at the same time facing financial uncertainty on their return to civilian life. They are also sometimes bitter about their service experience. As a result, they are potentially vulnerable to recruitment to espionage or could even voluntarily seek out deliberate compromise for money or revenge (Flyer, 1986).

Biodata. One type of instrument that could prove useful in suitability screening of enlisted accessions for sensitive jobs is the biographical or biodata questionnaire. Personal history items have been demonstrated to predict a variety of complex

behaviors in the private sector (Owen, 1976; Cascio, 1982). Also, a recent DoD Security Commission (1985) endorsed increased research in the application of behavioral science tests as a part of the prescreening and investigation procedures used to approve security clearances.

The Department of Defense (DoD) has initiated a study to develop and validate a biodata questionnaire to be used as an adaptability screening instrument (Atwater, Walker, & Weaver, 1987). The resultant biodata questionnaire was the Armed Services Applicant Profile (ASAP). An ASAP validation study against a criterion of first year attrition is currently being completed by the Navy Personnel Research and Development Center in San Diego, California (Trent, in preparation). The ASAP could provide an ideal vehicle for examining the predictive validity of a biographical questionnaire as a personnel security screening device.

Approach

Existing data from an ongoing ASAP validation study was used to examine the usefulness of biodata items as predictors of two personnel security criteria for personnel entering sensitive jobs: (1) issue case status during the background investigation, and (2) unsuitability discharge. The present article reports the results using the issue case criterion while a future paper will report data on unsuitability discharges.

ASAP total sample. All military enlisted applicants in the continental United States for the 3-month period from December 1984 to February 1985 completed one of two parallel forms of the ASAP. Each form consisted of 130 biodata items developed in previous research (see Atwater et. al., 1987). The ongoing DoD study is currently tracking 55,675 of these applicants who were categorized as non-prior service accessions.

Background investigation sample. The current focus was on those enlisted personnel applying for sensitive jobs that required a background investigation. Since this study was initiated after the recruits had successfully passed the prescreening procedures used by each service, there was no way to retroactively identify those recruits who were prescreened out. The result was that there was a restriction of range on the issue case criterion. Selective data on all background investigations is maintained on the Defense Index of Investigations (DCII). In order to identify ASAP accessions with background investigations, ASAP files were merged with the DCII files yielding 3,275 enlisted personnel who had both ASAP data and a completed BI. A breakout of this sample by the different services is presented in Table 1.

Criterion: Background investigation issue cases. During the conduct of BIs by the Defense Investigative Service (DIS), certain cases are categorized as issue cases. This indicates

that there may be adverse information in the person's background that reflects on that person's trustworthiness or reliability and thus or his or her qualifications to hold a high level security clearance. During the merging of the ASAP and DCII files, 12.1 percent of the final sample of 3,257 were categorized by DIS as issue cases. This rate varied from a high of 17.8 percent of the Navy BIs to 7.6 percent of the Marine BIs.

TABLE 1

Background Investigation Sample by Service

Service	Total Accessions N	Accessions with Background Investigations	
		N	%
Army	25,136	882	3.5
Navy	11,963	1,011	8.5
Air Force	12,350	1,259	10.2
Marines	6,226	1.05	1.7
Total	55,675	3,257	5.9

While issue case status reflects potential security unreliability, it also has been shown to predict subsequent unsuitability discharge form the military service Flyer (in press). Thus, while issue case status served as an interim personnel security criterion for the current study, it appears to also be linked to the key criterion of unsuitability discharge.

Scoring key development. Given the small sample size in the low criterion group (i.e., those personnel with issue case background investigations), and in order to double the N, 84 common items on both form 1 and form 2 of the ASAP were treated as a single shortened questionnaire. Approximately 52 percent (N = 1573) of the background investigation accessions were randomly assigned to a scoring-key development sample; the remainder (N = 1578) were assigned to a holdout sample for cross-validation purposes. The ASAP predictor score was developed using a weighted application method (see Crawford & Trent, 1987).

Results

Table 2 presents correlations (between ASAP scores and the issue case criterion) from the key development and cross-validation samples. The corrected cross-validation biserial correlations were all significant and were identical for Key A and Key B (r = .36) and slightly lower for Key C (r = .34).

Since Key B, based on five percent differences, yielded an identical cross-validity as the one percent key, yet required only 41 scored items, it was used to further examine the relationship between ASAP and issue case rates.

TABLE 2

ASAP Validities for Different Scoring Keys

Scoring Key % D	Items Scored	Item Responses	Key Development Sample ($\underline{N}$ = 1679)		Cross-Validation Sample ($\underline{N}$ = 1578)	
			P_{bis}	r_{bis}	P_{bis}	r_{bis}
A (1%)	83	250	.27	.44	.22	.36
B (5%)	41	62	.24	.39	.22	.36
C (10%)	8	10	.18	.29	.21	.34

Note: All point biserial and biserial correlations are significant ($\underline{p}$<.001).

Figure 1 shows that (in the cross-validation group) almost 24 percent of those individuals having the lowest 20 percent of the ASAP scores had issue case background investigations. Overall, the higher the ASAP quintile, the lower the rate of issues cases. Interestingly, approximately 42 percent of the issue case personnel scored in the lowest 20 percent quintile on the ASAP. In other words, if we eliminated the lowest 20 percent of the ASAP scorers, about 42 percent of those individuals with significant derogatory background information would be eliminated.

Figure 1. Issue Case Rate
by ASAP Score Quintile
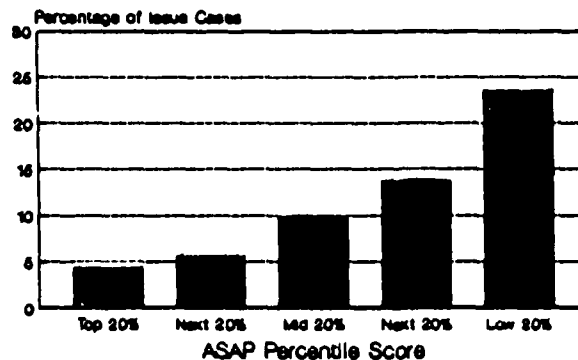
Different content categories of biodata items were also examined to determine which clusters showed the greatest discrimination between issue case and non-issue case personnel. These results indicated that issue case personnel were more likely to have used tobacco, had problems in school and on jobs, and committed a number of minor misbehaviors before entering the service (see Crawford & Trent, 1987, for more detail).

Conclusions

The results supported the potential use of ASAP type instruments as prescreens for personnel entering sensitive jobs. Additional analyses examining the relationship between ASAP scores and unsuitability discharges should further assess the usefulness of ASAP in security screening. However, Crawford and Trent (1987) identified a number of caveats for the current study that suggest that future issue case studies may have to be conducted using a more refined criterion. Also, the operational implications of using prescreening instruments need to be more clearly delineated. Nonetheless, it appears that a biodata test such as ASAP could potentially play an important role as a personnel security prescreening measure. Additional uses of an ASAP score could be as (1) a flag for conducting a more indepth BI, (2) an additional data element in the adjudication process, and (3) a flag for conducting a more thorough continuing evaluation if a security clearance is actually granted.

References

Atwater, D. C., Walker, C. B., & Weaver, C. N. (1987). Adaptability screening: Development of the Armed Services Applicant Profile (ASAP). San Diego: NPRDC. Unpublished report of the Adaptability Screening Group.

Cascio, W. F. (1982). Applied psychology in personnel management (2nd ed.). Reston, VA: Reston.

Crawford, K. S., & Trent, T. (1987). Personnel Security Prescreening: An application of the Armed Services Applicant Profile (ASAP) (PERS-TR-87-003). Monterey, CA: PERSEREC.

Department of Defense Security Review Commission. (1985). Keeping the nations secrets: A report to the Secretary of Defense by the commission to review DoD security policies and practices. Washington D.C.: Office of the Secretary of Defense.

Flyer, E. S. (1986). Personnel security research - prescreening and background investigations. Alexandria, VA: HumRRO.

Flyer, E. S. (in press). Characteristics and behavior of recruits assigned to highly sensitive positions. Alexandria, VA: HumRRO.

Owen, W. A. (1976). Background data. In M. D. Dunnette (Ed.), Handbook of industrial and organizational psychology (pp. 609-644). Chicago: Rand McNally.

Trent, T. (in preparation). Joint services adaptability screening: Initial validation of the Armed Services Applicant Profile (ASAP) (NPRDC TR). San Diego: NPRDC.

CONSEQUENCES OF PERCEIVED SELF
MOTION IN VISUAL DISPLAYS

PANEL DISCUSSION

Daniel W. Gower
U.S. Army Aeromedical Research Laboratory
A pilot's view of visual displays representing self motion.

Lawrence J. Hettinger
Essex Corporation
Visually-induced motion sickness.

Herschel W. Leibowitz
Pennsylvania State Univeristy
Sensory issues in the design of displays representing self motion.

Rik Warren
Armstrong Aeromedical Resarch Laboratory
Future cockpit displays.

Robert S. Kennedy
Essex Corporation
Adaptation to perceptual rearrangement and visual displays.

Sheldon Ebenholtz
State University of New York
Oculomotor issues in visual displays representing self motion.

CONSEQUENCES OF PERCEIVED SELF MOTION IN VISUAL DISPLAYS

PANEL DISCUSSION

Historic trends in flight simulation engineering technology indicate a commitment to developing the highest possible fidelity, and most realistic representations of the visual flight environment. The emphasis on enhanced realism in simulation is based on logical, popularly held, but essentially untested, assumptions concerning the degree of perceptual salience needed for training aviator skills in simulators, i.e., that greater realism leads to more effective training. Although there is as yet no definitive scientific data base, recent survey findings suggest that heightened realism might, ironically, lead to diminished training effectiveness because of the presence of undesired physiologic side-effects. These can include visual flashbacks, disequilibrium, eyestrain, disorientation, and nausea, which, in some cases, can persist for several hours following exposure to the simulation.

Emerging technologies for flight simulator design employ wide field-of-view displays with highly detailed computer-generated imagery to enhance a sense of realism which may result in the compelling visual illusion of self motion referred to as vection. Although the illusion of self-motion may be seen as desirable in terms of enhancing simulator realism, experimental evidence indicates that it may also lead to the disturbances mentioned above. This apparently negative influence of illusory self-motion appears to be due to recruitment of vestibular activity as a result of visual stimulation of a specific type, i.e., wide field of view, high spatial frequency.

Cockpits and related weapons systems of the future will be increasingly dependent on virtual image electronic displays. Many of these, as presently conceived, will entail wide field-of-view "wrap around" displays which could carry the "simulator sickness" syndrome into the operational environment. It is important to examine displays in the context of illusory motion and the ability of individuals to tolerate prolonged exposure to such an environment.

The increased occurrence of simulator sickness threatens the long-term utility of ground-based flight trainers as integral components in military and civilian flight training. Distrust and apprehension may develop among users of particularly troublesome simulators, limiting their training effectiveness. It may also be necessary in some cases to restrict post-simulator flight activities of users who experience sufficiently profound symptoms of sickness and disorientation, thereby diminishing their operational readiness. There is also a possibility that pilot trainees may adopt perceptual-motor strategies to avoid sickness in the simulator that will result in poor transfer of training to the aircraft. For example, in air combat maneuvering simulator flights, operators may learn to restrict head movements in order to avoid the onset of nausea produced by pseudo-Coriolis stimulation (Dichgans & Brandt, 1973). HEad movement restriction in actual combat would reduce the fiele of view and lessen one's chances of detecting an opponent, and thereby impact operational effectiveness.

92

It has been known for some time that symptoms of motion sickness can occur in the presence of visual stimulation alone, i.e., in the absence of any physical displacement (Lackner & Graybiel, 1979). Occurrences of illness in response to motion picture presentations of movement patterns in Cinerama (Benfari, 1964) and other wide field-of-view displays (Parker, 1971) have been reported. Lestienne, Soechting, & Berthoz (1977) reported that observers exposed to visually-specified, large-field patterns of high velocity, linear motion often reported intense, disturbing sensations. Three subjects out of thirty (10%) in their study were so disturbed that they fainted. The common element among these situations is the powerful, illusory sensation of self motion experienced by observers.

The neurological basis of visual influences on vestibular function has been the focus of a great deal of study in recent years (e.g., Precht, 1979). The neural interrelationships between the visual and vestibular systems, primarily through the vestibular nuclei and vestibular cerebellum, are integrated to the extent that the pickup of information by one quite often leads to perceptual consequences for the other. Therefore, it is generally more useful, particularly when considering issues of motion and simulator sickness, to conceptualize visual and vestibular proprioception as manifestations of an integrated perceptual system (Gibson, 1966).

Simulations of in-flight visual motion patterns vary in the extent to which they elicit illusory sensations of self motion. Some provide veridical representations of optical flow patterns characteristic of flight that do not lead to the illusion of self motion, while others appear to give rise to compelling experiences of self motion (Warren, 1982).

The distinction between the experience of illusory self motion, as opposed to the perception of a motion display depicting self motion with no concomitant experience of displacement, may be one of the keys to understanding the underlying causes of simulator sickness. Two possibly orthogonal aspects of visual motion perception might constitute a form of sensory mismatch that could lead to the sickness, particularly in fixed base flight simulators. Patterns of visually-specified motion, as presented in simulators, may give rise to two simultaneously perceived, but qualitatively distinct, percepts related to perceived velocity. On the one hand the observer perceives himself as stationary in space, but in a vehicle which is moving over the earth at the same rate of speed, due to characteristics of the flow pattern presented. On the other hand the observer perceives the displayed motion pattern as representing a stable environment and himself as moving. The first form is analogous to object motion perception in that although the observer may clearly comprehend that the depicted optical flow pattern accurately reproduces one that accompanies motion, no compelling illusion of self-motion is actually presented. The second form involves visually-specified illusory self-motion, or vection.

Vection may entail a significant vestibular element while the perception of a display representing self motion with no illusory displacement may not. A number of studies (e.g., Held, Dichgans, & Bauer, 1975; Mauritz, Dichgans, & Hufschmidt, 1977) have demonstrated large effects of rotating visual displays on postural sway. Observers in these studies continually readjusted their

stance to compensate for visually-specified displacements of the gravito-inertial upright. Lestienne et al. (1977) reported similar effects with patterns representing linear motion.

These studies indicate a strong influence of visual motion patterns on vestibular function. Their relevance for the design of flight simulators and other future visual displays lies in the demonstration that representations of motion patterns may often exert strong influences on the vestibular system. Those displays which produce strong vestibular effects may be the most bothersome in terms of producing simulator sickness. In order to alleviate simul :or sickness and possible aftereffects of visual-vestibular stimulation a converging operation (Garner, Hake, & Eriksen, 1956) whereby we: (a) investigate the training utility of displays which do not produce illusory self motion, and (b) identify those properties of vection-producing displays which lead to disturbances so they can be eliminated.

While it is clear that the pattern of motion depicted on a visual display is important for the perception of translatory self-motion and the onset of simulator sickness, there are no satisfactory or complete descriptions of the relevant characteristics of optical flow patterns in operational simulators. Although we know how to produce these patterns, their description in relation to the effects of visually-specified motion on observers is only beginning to be attempted (Warren, 1982; Owen, 1982, 1983). Researchers disagree on the requirements of visual displays for producing illusory self motion. Indeed, the phenomenological nature of self-reports of vection renders it problematic as an experimental metric.

Investigations of visual-vestibular disturbances carried out to date indicate that visual displays that impact on vestibular function are likely to be involved in the etiology of simulator sickness. The identification of the parameters of these displays will provide useful engineering guidelines to prevent the occurrence of this malady.

REFERENCES

Benfari, R. C. (1964). Perceptual vertigo: A dimensional study. Perceptual and Motor Skills, 18, 633-639.

Dichgans, J., & Brandt, T. (1973). Optokinetic motion sickness as pseudo-Coriolis effects induced by moving visual stimuli. Acta Otolaryngolica, 76, 339-348.

Garner, W. R., Hake, H. W., & Eriksen, C. W. (1956). Operationism and the concept of perception. Psychological Review, 63, 317-329.

Gibson, J. J. (1966). The senses considered as perceptual systems. Boston: Houghton-Mifflin.

Held, R., Dichgans, J., & Bauer, J. (1975). Characteristics of moving visual scenes influencing spatial orientation. Vision Research, 15, 357-365.

Lackner, J. R., & Graybiel, A. (1979). Some influences of vision on susceptibility to motion sickness. <u>Aviation, Space, and Environmental Medicine</u>, <u>50</u>, 1122-1125.

Lestienne, F., Soechting, J., & Berthoz, A. (1977). Postural readjustments induced by linear motion of visual scenes. <u>Experimental Brain Research</u>, <u>28</u>, 363-384.

Mauritz, K. H., Dichgans, J., & Hufschmidt, A. (1977). The angle of visual roll determines displacement of subjective visual vertical. <u>Perception & Psychophysics</u>, <u>22</u>, 557-562.

Owen, D. H. (Ed.). (1982). <u>Optical flow and texture variables useful in simulating self motion: I</u> (Interim Tech. Rep. for Grant No. AFOSR-81-0078). Columbus, OH: Ohio State University, Department of Psychology, Aviation Psychology Laboratory.

Owen, D. H. (Ed.). (1983). <u>Optical flow and texture variables useful in simulating self motion: II</u> (Final Tech. Rep. for Grant No. AFOSR-81-0078). Columbus, OH: Ohio State University, Department of Psychology, Aviation Psychology Laboratory.

Parker, D. M. (1971). A psychophysiclogical test for motion sickness susceptibility. <u>Journal of General Psychology</u>, <u>85</u>, 87-92.

Precht, W. (1979). Vestibular mechanisms. <u>Annual Review of Neurosciences</u>, <u>2</u>, 265-289.

Warren, R. (1982). <u>Optical transformations during movement: Review of the optical concomitants of egomotion</u>. (AFOSR 81-0108). Columbus, OH: The Ohio State University, Department of Psychology, Aviation Psychology Laboratory.

HUMAN PERFORMANCE ISSUES IN VISUAL DISPLAYS

Historic trends in flight simulation engineering technology indicate a commitment to developing the highest possible fidelity, and most realistic representations of the visual flight environment. The emphasis on enhanced realism in simulation is based on logical, popularly held, but essentially untested, assumptions concerning the degree of perceptual salience needed for training aviator skills in simulators, i.e., that greater realism leads to more effective training. Although there is as yet no definitive scientific data base, recent survey findings suggest that heightened realism might, ironically, lead to diminished training effectiveness because of the presence of undesired physiologic side-effects. These can include visual flashbacks, disequilibrium, eyestrain, disorientation, and nausea, which, in some cases, can persist for several hours following exposure to the simulation.

Emerging technologies for flight simulator design employ wide field-of-view displays with highly detailed computer-generated imagery to enhance the compelling visual illusion of self motion referred to as vection. Although the illusion of self-motion may be seen as desirable in terms of enhancing simulator realism, experimental evidence indicates that it may also lead to the disturbances mentioned above. This apparently negative influence of illusory self-motion appears to be due to recruitment of vestibular activity as a result of visual stimulation of a specific type, i.e., wide field of view, high spatial frequency.

Cockpits and related weapons systems of the future will be increasingly dependent on virtual image electronic displays. Many of these, as presently conceived, will entail wide field-of-view "wrap around" displays which could carry the "simulator sickness" syndrome into the operational environment. It is important to examine displays in the context of illusory motion and the ability of individuals to tolerate prolonged exposure to such an environment.

The following individuals are proposed as speakers and panel discussants for a session intended to present scientific data and implications for the design of simulator visual displays, illusory self-motion, and human performance. Each individual is well known in the area of applied visual problems of flight simulation and each has carried out related research on these issues for agencies of the Department of Defense.

Panel Session
Multi-service Stress Prevention and Intervention Programs

A. David Mangelsdorff, PhD., M.P.H.
Health Services Command
Fort Sam Houston, Texas 78234-6060

Stress and its potential effects have become topics of considerable interest to the Armed Services. The dynamics of being a victim of a crime, or of a terrorist activity, or involvement an accident, a disaster, or combat can have long-term mental health consequences. A variety of professions (such as being a member of an air crew or of being a military leader) have unique occupational stresses associated with the work environments. Providers of health, security, and emergency services are potential casualties because of the nature of the work in which they are involved. Being a member of particular military groups in operational settings (such as basic trainees or academy cadets) can be extremely stressful. Unit leadership is critical to enhancing the climate and cohesion of the members within a unit, whether in the active forces or in the Reserves. This panel will examine the nature of some of these stresses, the prevention and intervention programs that are being used to deal with the stressors, and the possible consequences of these programs.

The papers in this multi-service panel include:

"Victims of Crime: Epidemiological and Treatment Considerations"
Connie L. Best, PhD.; Medical University of South Carolina

"Psychological Research on Torturers: An Overview
Richard Bloom, PhD., ABPP; Organization of the Joint Chiefs of Staff

"Persons, Environments, and the Management of Stress"
Paul Bartone, PhD.; Walter Reed Army Institute of Research

"Subordinate Feedback to Unit Commanders"
Terrence D. Fullerton, PhD; U.S. Military Academy West Point

"A Model Soldier Sustainment on the Battlefield: Soldier
Commitment and Small Unit Leadership." James Griffith, PhD.;
Westat

"USAF Academy Cadet Stress-Induced Immune Suppression"
Richard Meehan, M.D.; David Jenkins, PhD; University of Texas Medical
Branch at Galveston. Jeffrey S. Austin, PhD and Robert C. Ginnett,
PhD; USAF Academy

"Stress Inoculation in Operational Settings"
Thomas P. O'Hearn, PhD; Front Range Institute

"Who Is At Risk? Health Care Providers are Potential Casualties Too"
A. David Mangelsdorff, PhD, P.P.H.; Health Services Command

Victims of Crime: Epidemiological and Treatment Considerations

Connie L. Best, Ph.D.
Medical University of South Carolina
Charleston, South Carolina 29425

Abstract

The prevalence of victims of violent crime among adult women is higher than previously believed. What strategies can the victims learn to cope with the mental health consequences of their traumatic stress? What are the assessment, treatment, and evaluation imnplications of these findings?

Recent research shows that the prevalence of sexual assault and other violent crimes among adult women is higher than previously thought, and that many victims experience long-term mental health problems such as fear, anxiety, and depression. Thus, mental health professionals need to learn more about how to treat victims and their problems. This paper will present the culmination of more than a decade of studies of post-assault sequelae and treatment efficacy.

Specifically, data from five research grants will be described. The first is a National Institute of Mental Health (NIMH)-funded longitudinal study comparing sexual assault victims with matched control subjects. Assessment of victims and nonvictims were made at one month, three months, six months, one year, eighteen months, two years, three years, and four years post assault. The results yielded information regarding: type, severity, and timing of psychological responses to sexual assault, and also provided evidence as to the long-term nature of assault-related problems. Another study funded by the National Institute of Justice (NIJ) assessed the lifetime prevalence of violent crimes against women. Data from a representative random sample of 2004 women in the Charleston, South Carolina area involving eight major categories of crimes revealed that nearly one quarter of the adult women have been victims of a violent crime. Of those women who were victims of a completed sexual assault, one in five had subsequently made a suicide attempt. In a second NIJ study, a percentage of the women in the above study were recontacted and assessed for mental health consequences of criminal victimizations as based on a modified version of the Diagnostic Interview Survey of the DSM III psychiatric classification system. The results indicated one-fourth of the women who were crime victims developed a post-traumatic stress disorder (PTSD) at some point after the crime, and at the time of assessment, 7.5% of the women were currently suffering from PTSD.

The fourth study concerns the efficacy of a specific treatment developed at the Crime Victims Research and Treatment Center (CVC) of the Medical University of South Carolina with that of a pharmacological treatment for rape-related fears and PTSD. The fifth study examines methods and strategies for reducing anxiety and stress for victims of crime who are involved in court proceedings.

In addition to presentation of research findings, this paper will describe treatment interventions developed for fear and anxiety association with criminal victimizations. Based on a learning theory model, a modified stress inoculation training treatment package has been developed and implemented for post-traumatic stress disorders. A description of assessment, treatment, and evaluation strategies will be elaborated.

Psychological Research on Torturers: An Overview

Richard W. Bloom, Ph.D., ABPP
Directorate of Operations
Organization of the Joint Chiefs of Staff

Abstract

This paper provides an overview of psychological research on
torturers in political and military organizations. Research
methods have included interviews, psychometric testing, and
analog studies. Each method has significant limitations.
Research findings comprise proposed psychological mechanisms for
engaging in torture. These findings may reflect gross
interpretive distortions of raw data by researchers. Even
improved research may not support primary prevention programs for
military and political organizations, all of which are liable to
employ torture.

This paper describes (a) research methods used to study torturers in
political and military organizations, (b) research findings on psychological
mechanisms posited for engaging in torture and (c) problems with both methods
and findings.

RESEARCH METHODS. Psychologists have interviewed and tested torturers
both before and after the latter have renounced torture (Gibson, 1986).
Obtaining access to torturers can be dangerous, especially if one publically
espouses disapproval of torture or the goals of the authorities that employ
torturers. To pose as a *supporter of torture* facilitates access, but poses
significant ethical and legal questions concerning becoming an accessory. At
least several psychologists may have provided industrial/organizational
consultation services for torturers in exchange for interview material.

Psychologists have analyzed interview data about torturers including
verbal reports from torturers' acquaintances and victims (Bloche, 1986). These
data often have been collected and subjected to preliminary analyses by
journalists, political activists, and human rights advocates.

Psychologists have abstracted inferences about torturers from laboratory
and field studies. These studies are not about torture, but topics such as
compliance, conformity, aggression, and deception (cf. Zimbardo, 1969). Some
psychologists view these topics as elucidating mechanisms assumed salient in
engaging In torture.

RESEARCH FINDINGS. From raw data, psychologists have posited a number of
mechanisms, mostly intrapsychic in nature, to "explain" why torturers torture.
Almost always these mechanisms are advanced as antecedent and causally related
to engaging in torture.

Devaluation. The torturer values the victim less than other people
are valued (Kelman, 1986). The victim is perceived as less human, as animal
(in a sense not embraced by animal liberation members), as not worthy of
living, of no value alive or dead. As the distance between a potential victim

100

and a potential torturer on the phylogenetic scale increases, so does the probability of torture occurring.

Moral Disengagement. Most of the time the torturer is no different than anyone else. The torturer shares similar moral standards with others in congruent cultures. He or she is engaged with these standards. However, the torturer is able to disengage from these standards so as to torture. The standards incipient to, during, and immediately after one engages in torture are radically different from those engaged in at other times (Lifton, 1986).

Projection. An individual cannot consciously accept certain personal characteristics which rightly or wrongly could be used as self-ascriptions. Accepting these characteristics as self-ascriptions would result in significant reductions of self-esteem, even ego integrity. So, the individual unconsciously transfers self-ascriptions to the perceived characteristics of other people. These characteristics are then consciously perceived in others, often in an exaggerated fashion. Some of them become potential and real victims of torture through devaluation or through fear of their projected characteristics (Valladares, 1986).

Ideology. An individual's beliefs on how power is and should be distributed throughout the world can affect the probability of engaging in torture. First, the ends of ideology can be so intensely embraced as to legitimize the use of any means, including torture. Second, the content of ideology can embrace torture not only as a means, but as an end itself (Cheng, 1987).

Situational Pull. Individuals that otherwise might lead unremarkable lives will engage in torture in remarkable situations. Paradoxically, these remarkable situations may be perceived by these individuals as quite unremarkable, hence Arendt's term "banality of evil". A related phenomenon is the effect that demand characteristics of authority will have on an individual's behavioral repertoire. As Milgram (1974) has shown, these demand characteristics affect a significant minority of otherwise unremarkable people.

Compliance and Dependency. With some individuals, situational pull is either not necessary or pertains to such a broad range of situations as to make the concept of "pull" meaningless. Torture can occur mainly to satisfy strong needs to comply, to please others, or to ward off the feared loss of reinforcers that can be available from others. These reinforcers are often integral to one's sense of self-integrity and self-worth (Gao, 1987).

Self-Maintenance. Torture can not only provide a homeostatic function that helps balance competing needs but also contributes to the very sense of self, especially self-identity. Perhaps I think, therefore I am. Perhaps I torture, therefore I know who I am (Levi, 1947).

Sadism. Few torturers seem to be sadists. Instead torture seems to be an equal opportunity employer, for people like you and me (Staub, 1986).

Other Mechanisms. There are a number of mechanisms that are posited for engaging in any behavior that violates publicly proclaimed social norms. These include (1) blaming the victim for the necessity for torture, (2) minimizing, ignoring, or misconstruing the consequences of torture, (3) displacement or diffusion of responsibility, (4) moral justification, and (5) euphemistic labeling (see Bandura, 1986).

RESEARCH PROBLEMS.

Methods. (1) The interpretation of interview data without an actuarial foundation usually is the least reliable and valid form of assessment. (2) The most significant problem with test data on torturers is subject self-selection as well as the nonrandom effects responsible for those who have been identified, captured, and convicted. (3) Those psychologists who attempt to exploit studies not about torture assume a priori what they conclude a posteriori. Moreover, the subjects of these studies usually are white, middle-class college students whose torturous angst have tortuous relevance to the real thing.

Findings. (1) Psychologists seem to have taken the stance that torture represents good people going bad or going wrong. The same data also could suggest that people usually are bad or going wrong and only occasionally are good. The interpretive bias of most psychologists certainly reflect Rousseau more than Hobbes, as well as social perceptual styles suggesting the Pollyana hypothesis. During the Twentieth century, the folks responsible for the use of poison gas in World War I, enforced Soviet collectivization and industrialization during the 1930s, the Holocaust, genocide in Cambodia, and good, old-fashioned atrocities give this hypothesis a strong run for its money.

(2) Psychologists have focused on proposed intrapsychic mechanisms to the relative exclusion of environmental factors. In this respect, researchers have been guilty of the fundamental attribution error. Here they embrace a negative value towards torturers and, thus, are more prone to attribute torturers' behaviors to dispositional phenomena.

(3) Many of the proposed mechanisms related to torture are neither necessary nor sufficient for torture to occur. They may be epiphenomenal, modifying, moderating, or otherwise correlative, not causative.

(4) Researchers' moral agendas usually entail castigating and preventing torture. Yet these researchers rarely note that the psychological mechanisms posited to "explain" torture often are adaptive for all people. This, in turn, poses significant problems for preventing torture.

THE VALUE OF RESEARCH ON TORTURERS. Even if psychologists could elucidate valid mechanisms underlying the manifestation of torture behavior, the probability of this knowledge contributing to preventing torture would be minimal. This is because many behaviors, especially sexual and aggressive ones, seem to be overdetermined and dependent on psychological dynamics. As effective cognitive-behavior and situation modification programs were implemented, new mechanisms would appear to maintain old behaviors. As with other behaviors such as espionage, prostitution, and murder, at best we can hope for management, not for eradication. Thus, psychologists need not fear a complete solution obviating a need for further research and practice-- ambivalence towards carrying out superior research is not warranted.

References

American Psychological Association.   (1987, Draft). Report of
     Subcommittee on the Abuse of Psychological Knowledge for Purposes of
     Inflicting Torture. Washington, DC: Author.
Bandura, A.   (1986, March). Mechanisms of moral disengagement.
     In W. Reich (Chair), The Psychology of Terrorism: Behaviors, world views,
     states of mind. Conference, International Security Studies Program,
     Woodrow Wilson International Center for Scholars, Washington, DC.
Bandura, A. (1986).   Social foundations of thought and action: A
     social cognitive theory. Englewood Cliffs, NJ: Prentice-Hall.
Bloche, M.G.   (1986). Uruguay's military physicians: Cogs in a system of
     state terror. Journal of the American Medical Association,
     255(20), 2788-2793.
Cheng, N.   (1987). Life and death in Shangai. Grove Press.
Cordes, C.   (1984). Psychologist said to be architect of abuse.
     The APA Monitor, December.
Gao, Y. (1987). Born red: A chronicle of the cultural revolution.
     Stanford University Press.
Gibson, J. T.   (1987, August). Psychology of torture: Etiology, dynamics,
     prevention. In R. W. Bloom (Chair), Psychology and National
     Security Affairs. American Psychological Association Preconvention
     Workshop, Division of Military Psychology, New York.
International Association on the Political Use of Psychiatry   (IAPUP)
     Information Bulletin No. 7. October 1983.
Kelman, H. C. (1973). Violence without moral restraint: Reflections on
     the dehumanization of victims and victimizers. Journal of Social Issues,
     29, 25-61.
Levi, P. (1947). Survival in Auschwitz. Summit Books.
Levi, P. (1961). The reawakening. Summit Books.
Lifton, R. (1986). The Nazi doctors: Medical killing and the psychology of
     genocide. Basic.
Medical ethics and the issue of torture. (1986). Journal of the American
     Medical Association, 255(20), 2798-2799.
Milgram, S. (1974). Obedience to authority: An experimental view. New York:
     Harper & Row.
Puddington, A. (1987). Pol Pot in retrospect. Commentary, 49-54.
Staub, E. (1986, August). The psychology of torture and torturers.
     In P. Suedfeld (Chair), Psychology and torture, American
     Psychological Association Annual Convention, New York.
Stover, E. & Nightingale, E. O.   (1985). The breaking of bodies and
     minds. N. Y.: Freeman.
Valladares, A. (1986). Against all hope: The prison memoirs of Armando
     Valladares. Alfred A. Knopf.
Zimbardo, P.G.   (1969). The human choice: Individuation, reason, and order
     versus deindividuation, impulse, and chaos. In W.J. Arnold &
     D. Levine (Eds.), Nebraska symposium on motivation, 1969, Lincoln:
     University of Nebraska Press.

Disclaimer

This paper does not necessarily reflect the views of the Organization of
the Joint Chiefs of Staff or the Department of Defense.

Persons, Environments, and the Management of Stress

Paul Bartone, Ph.D.
Department of Military Psychiatry
Walter Reed Army Institute of Research
Washington, D.C. 20307

Abstract

Environmental stress can have significant effects on health, well-being, and performance. How can we account for the variations in individual differences in response to stress? What are the characteristics about work environments that make them stressful? What strategies can be developed to make individuals more hardy to coping with stress?

The evidence from behavioral and medical research is now clear: environmental stress can have quite harmful effects on health, well-being, and performance. Whether the stress comes in the form of major, discrete life-events or trauma, or in the form of chronic occupational stress or "day-to-day hassles," the impact on individuals can be severe. However, the evidence is just as clear that people vary tremendously in how they respond to or manage stress in their lives. What accounts for these variations? According to Horowitz (1986), this is the important issue around which research on stress and illness should now be focused. Such research has important practical implications. If we can answer the question as to individual differences in response to stress, we may gain some important clues as to what the effective stress management program should contain, and how stress-related illness or injury can be reduced or prevented.

Recent research conducted at the Walter Reed Army Institute of Research has found both social and personality variables to be significant determiners of resistance to the ill effects of stress. Our studies show that both chronic soldier occupational stress and traumatic stress related to disaster are less damaging to individuals who possess a personality style described best as "hardiness," and who report particular kinds of social support (Bartone and Hoover, 1987; Bartone, Ursano, Saczynski, & Ingraham, 1987). Personality hardiness has also appeared as an important moderator of the stressful impact of providing assistance and emotional support to bereaved relatives following a disaster, and of related role ambiguity or confusion (Bartone, 1987). The support of one's superiors for the role of assistance provider also emerged as an important stress-resistance resource. Managers or leaders in organizations can apparently exert more general influence on how their employees manage stress as well. Schneider (1986) recently showed that managers in small business settings establish psychological work environments that are more and less conducive to the generation of stress-related symptoms among employees (Schneider, 1986).

Taken together, these studies indicate that both characteristics of situations and of persons affect outcomes of stressful encounters. Effective stress management programs should thus be concerned not only with what tools or resources the individual has available to manage stress, but also with aspects of the environment that cause (or exacerbate) stress. The recent experience of several urban transit authorities shows that when stress-inducing aspects of situations that are "controllable" or modifiable are ignored, then employee participation in stress management programs suffers (Maddi, 1987). The credibility and success of stress management programs thus hinge on a recognition of, and a proactive stance toward controllable sources of environmental stress.

Three specific implications for stress control intervention strategies might be drawn from these studies: (1) when aspects of situation causing "stress" can reasonably be eliminated or reduced, they should be; (2) the "hardy" style and behavior pattern should be taught and reinforced to the extent this is possible; and (3) situations (including leadership climate) should be structured so as to encourage the "hardy" approach to dealing with stress.

References

Bartone, P. (1987). Boundary crossers: The role of Army family assistance officers in the Gander Crash. Paper presented at the Biennial Conference of the Inter-University Seminar on Armed Forces and Society, Chicago, Illinois.

Bartone, P. and Hoover, E.C. (1987). Occupational stress, health, and general well being among soldiers. Paper presented at the 95th Annual Convention of the American Psychological Association, New York, New York.

Bartone, P., Ursano, R., Saczynski, K. & Ingraham, L. Impact of a military air disaster on the health of assistance workers: A prospective study. Paper presented at the 95th Annual Convention of the American Psychological Association, New York, New York.

Horowitz, M. (1986). Disaster stress studies: Conclusions. In J. Shore (Ed.), Disaster Stress Studies: New Methods and Findings. Washington, D.C.: American Psychiatric Press.

Maddi, S.R. (1987) Commitment, control and challenge; and Strategies for the future. Papers presented at the 1987 Conference on Transit Industry Stress, University of Wisconsin, Milwaukee, Wisconsin.

Schneider, M.G. (1986). The boss and organizational stress. Unpublished doctoral dissertation, Department of Behavioral Sciences, University of Chicago.

Subordinate Feedback to Unit Commanders

Terrence D. Fullerton, Ph.D.
U.S. Military Academy
West Point, New York 10996

Abstract

What will be the effect of providing subordinate feedback to unit commanders on the subordinates' perception of the cohesion and command climate in the unit? How does command climate affect unit performance? How do commanders feel about receiving specific feedback?

This paper is part of an ongoing study attempting to provide to unit commanders subordinate feedback on cohesion and command climate in their unit. Up to now, the focus of many agencies collecting soldiers' perceptions of these areas has been to glean from the data lessons to be provided to the Army as a whole. This orientation has prevented in-depth feedback to specific units, e.g., to prevent the contamination of future results since those units are still part of an ongoing study. Having units not part of that long term project complete the same instrument and subsequently provide feedback to those commanders gives a perspective on the usefulness of subordinate feedback to the commanders and problems associated with either providing the feedback or the commander's attempts at intervention.

AirLand Battle doctrine requires units whose soldiers are tightly bonded and units with a positive command climate. Army leaders have made organizational changes, e.g., Unit Manning System with its COHORT (Cohesion, Operational Readiness and Training) Program and the Regimental System, to strengthen the bonds of soldiers with each other, the leader, and the unit. There are also calls for other measures to focus additional attention on improving command climate. Knowlton (1987) and the Strategic Studies Institute of the Army War College call for subordinate ratings of commanders.

Knowlton (1987) suggests that positive command climate is critical for the development of initiative and confidence in subordinate leaders. He cites evaluations at the National Training Center to show the criticality of command climate during combat-like situations. The "Professional Development of Officers Study" suggested that commanders control the unit climate and that officers need to demonstrate teaching and a caring leader style. Command climate is defined by Vaitkus (1987) as the shared perceptions of unit membership about the quality of leadership within their unit. The present and previous Army Chiefs of Staff have articulated the importance of caring leadership. Knowlton suggests reinforcing that goal by providing subordinate evaluations directly to MILPERCEN as part of the total view of that commander's performance. Thus a commander faces the choice of being concerned and caring of subordinates and paying attention to command climate or adversely affecting his own evaluation.

As mentioned above, another approach to improve command climate is to provide feedback from his subordinates on unit cohesion and unit climate to the commander via survey results. This would provide a unique viewpoint that the commander might otherwise not know. Results of a study focusing on the difference in perceptions of officers, NCOs, and enlisted soldiers of the cohesiveness of their unit reveal stark differences between officers and the other two groups (Vaitkus, 1987). A study examining subordinate feedback in an industrial setting suggests that while there are advantages and disadvantages to providing subordinate feedback, overall the organization, the superior, and the subordinate benefit. These results reveal that 88% of superiors getting subordinate feedback attempted to change the work environment and their behavior; 75% of the superiors wanted to continue to get feedback; and 25% of the subordinates saw lasting and major improvements in the behavior of their superiors (Hegarty, 1981). In a military setting, a form of subordinate feedback, a morale survey, is presently conducted on a regular basis in the Israeli Defence Force (Gal, 1986). In the IDF, results are provided by a psychologist to the commander of the unit whose soldier completed the IDF morale survey. This project attempts to model its feedback and interventions on those currently in use in the IDF.

Method

Subjects. Data was collected from four Ranger companies (n=300) for comparison to data being collected as part of an ongoing evaluation of the Unit Manning System. As part of this project the Department of Military Psychiatry, WRAIR evaluated every six months over 100 companies (n=9,000 soldiers). These soldiers provided anonymous responses.

The Instrument. The instrument was used as part of a larger questionnaire in the Unit Manning System study. The items came from an Americanization of the IDF survey given to the lead elements preparing for the invasion of Lebanon, items from surveys used in World War II, and items developed at WRAIR. It was designed to assess the six areas listed below.

1 Concerned Leadership and Vertical Cohesion. These 14 items focus on leader-soldier relationships, the leader's personal interest in his soldiers, and leader-soldier relationships and cooperation.

2 Horizontal Cohesion. These 18 items focus on the trust, teamwork, and cooperation among unit members, friendships, and who would help them with personal or debt problems.

3 Confidence in Company Level Leadership. Five items focus on soldiers' view of their NCOs' competency and how well their NCOs would lead in combat. Nine items focus on soldiers' views of their confidence in their officers' competency and how well their officers would lead in combat.

4 Confidence in Senior Level Leaders. Five items focus on soldiers' confidence in the tactics of the Batallion Commander through the Army General Staff.

107

5 Combat Readiness. Fifteen items focus on soldiers' confidence in unit readiness, their weapons, their level of training, and their willingness to trust their lives to the combat skill of their company.

6 Personal (Individual) Morale. Six items focus on their own personal morale, confidence in their own combat skills, and worth to the unit.

Results and Feedback

The results are currently being analyzed by factor analysis and between unit measures. Initial results indicate the unit getting feedback did quite well on many of the measures. Comparisons to the other units and how the commanders view the feedback will be presented.

References

Gal, R. (1986). Unit Morale: From a theoretical puzzle to an empirical illustration - An Israeli example. Journal of Applied Social Psychology, 16, 549-564.

Hegarty, W.H. (1981). Using subordinate ratings to elicit behavioral changes in supervisors. In B.M. Bass (editor) Handbook of Leadership: A Survey of Theory and Research. New York: The Free Press.

Ingraham, L. H. and Manning, F.J. (1981). Cohesion: Who needs it, what is it, and how do we get it to them? Military Review, 61, 2-12.

Knowlton, W.A. (1987). In rating the leaders, ask the led. Army, 21-25.

Manning, F.J. and Ingraham, L.H. (1983). An investigation into the value of unit cohesion during peacetime. (Report number NP-83-5). Washington, D.C.: Walter Reed Army Institute of Research.

Marlowe, D. (ed) (1985). New Manning System Field Evaluation Report number 1. Technical report. WAshington, D.C.: Walter Reed Army Institute of Research.

(1984). Professional Development of Officers Study.

Stouffer, S.A., Devinney, L.C., and Star, S.A., and Williams, R.M., (1949). Studies in Social Psychology in World War II: Volume 2. The American Soldier. Princeton, NJ: Princeton University Press.

Strategic Studies Institute (1985). Senior Field Grade Officer Motivation and Professionalism study. Carlisle Barracks, PA.: Army War College.

Vaitkus, M. (1987). Longitudinal analysis of cohesion measures in military units. Paper presented at the meeting of the 95th Annual Convention of the American Psychological Association, New York, New York, August.

A Model for Soldier Sustainment on the Battlefield:
Soldier Commitment and Small-Unit Leadership

James Griffith, Ph.D.
Westat, Inc.
Rockville, Maryland

Abstract

The purpose of this paper is to provide a preliminary model of the social-psychological ingredients affecting soldier sustainment on the battlefield. The model is a result of analysis of survey and group interview data, and a review of the pertinent literature. As group membership provides the necessary psychological motivation for soldiers to continue combat performance under high intensity combat conditions, individual soldier commitment to the military organization forms the basis of this model. The utility of this model is demonstrated by tying together concepts and constructs found in the relevant academic literature and past and current Army research. Specific data collection methods and an analysis plan are presented to examine interrelationships among the model's constituent constructs and their relationship to criterion measures. The model's implications for examining the deleterious effects of external stresses on organizational social structures that provide sustainment for soldiers during high intensity combat are also discussed.

The Model of Soldier Commitment

This paper describes a preliminary model of the social-psychological ingredients that sustain the soldier's membership in the military organization, and thereby, his effectiveness as a combatant. The underlying assumption of this model is that the individual soldier is better able to perform the combat role in military groups with which he strongly identifies rather than groups from which he feels alienated and alone. This assumption has strong intuitive appeal. In addition, past studies have shown that social structures of the military organization provide the necessary social and psychological elements for motivating the soldier to fight. Military groups provide the soldier a sense of purpose and global meaning, social identity, and support in coping with stressful circumstances (Kobassa & Pucetti, 1983; Marshall, 1978; Steiner & Neumann, 1978; Solomon, Mikulincer, & Hobfoll, 1986).

The extent to which the soldier identifies with the military organization, then, forms the cornerstone for a combat effective soldier (Shils & Janowitz, 1948). A review of the current literature pertinent to organizational commitment, cohesion and leadership showed two emergent aspects of commitment. The model depicts the mechanisms of commitment as affective commitment -- a sense of loyalty and allegiance -- and instrumental commitment -- sense of contractual obligation involving a mutual exchange of services between the individual and the organization. Commitment implicates referent groups, and the type of referent group becomes the second dimension of commitment. The primary referent groups (e.g., the squad, company) describe organizational subgroups in which soldiers have daily face-to-face contact with group members, whereas the secondary referent groups (e.g., the battalion, regiment) describe as organizational

subgroups with which the soldier has limited social interaction, or even, are abstract referents (e.g., the NCO and Officer Corps, the Army, country).

The model depicts small-unit leadership (company level and below) as the primary means for developing and maintaining soldier commitment. Many problems of the peacetime soldier can be resolved at the small-unit level, and their lack of resolution undermines the confidence and trust the individual soldier has in his(her) immediate leaders and the Army "system" in general, resulting in dissatisfaction, greater alienation, lower commitment and lower likelihood that the soldier will stand and fight. The small-unit leaders give organizational goals, values, and norm clarity and influence their acceptance at the primary group level. Group members' identification with primary group leaders often results in commitment of soldiers to the norms and values of the larger organization. To the soldier, the NCO and officer represent an extension of the Army as an institution and the extent to which these leaders embody the values of the larger organization, the greater the felt allegiance beyond the primary reference group. In this way, the small-unit leader provides the critical link between the primary and secondary referent groups and are invaluable in transmitting organizational culture and climate to lower organizational elements (Allport, 1933; Shils & Janowitz, 1948).

Such a typology is useful in elaborating on both the referent group and the type of commitment for important concepts found in existing military literature. For example, instrumental secondary commitment refers to Moskos' (1977) occupational orientation, whereas affective secondary commitment refers to his institutional orientation. These dimensions are also useful in specifying operational aspects of concepts found in past and current military studies conducted by the Army Research Institute, the Walter Reed Army Institute of Research and the Israeli Defense Force. For example, the terms "vertical and horizontal cohesion" refer to the referent group of the soldier's commitment. Vertical cohesion describes the extent to which soldiers are committed to their leaders in their primary and secondary groups. Horizontal cohesion describes the level of commitment among primary groups (e.g., lower-ranking soldiers).

Analysis Plan

A specific method to assess the model's constituent parts and empirically validate the model of soldier commitment is presented. An analysis plan to examine the interrelationships among the constituent parts in addition to how these relationships are moderated by unit characteristics (e.g., unit leadership) and external characteristics (e.g., stress) is presented. The appropriateness of performance indicators, both individual and collective, for evaluating the effects of soldier commitment on combat performance is discussed.

External Stress and Organizational Structure

The model emphasizes the individual soldier's sense of commitment to referent groups in the military organization. Further, it is argued that such membership and its quality provide a buffer against the dehabilitating effects of extreme stress. Such a perspective implies that it is not sufficient to only study the relationship between the soldier's sense of commitment and combat stress, but also the relationship between combat stress and the social structure of the military organization, especially those structures critical to the social psychological sustainment of the soldier. Sociological analyses of civilian disasters, for example, have shown that the natural disasters are more destructive

110

by their disruption of the community social structure, rather than the loss of human life and injury (Erikson, 1976).

References

Allport, F. H. (1933). Institutional behavior. Chapel Hill: University of North Carolina Press.

Erikson, K. T. (1976). Everything in its path: Destruction of the community in the Buffalo Creek Flood. NY: Simon & Schuster.

Kobassa, S. C., & Pucetti, M. C. (1983). Personality and social resources on stress resistance. Journal of Personality and Social Psychology, 45, 839-850.

Marshall, S. L. A. (1978). Men against fire: The problem of battle command in future war. Gloucester, MA: Peter Smith.

Moskos, C. C. (1977). From institution to occupation: Trends in military organizations. Armed Forces and Society, 4, 41-50.

Shils, E. A., & Janowitz, M. (1948). Cohesion and group disintegration in the Wehrmacht. Public Opinion Quarterly, Summer, 280-315.

Solomon, Z., Mikulincer, M., & Hobfoll, S. E. (1986). Effects of social support and battle intensity on loneliness and breakdown during combat. Journal of Personality and Social Psychology, 51, 1269-1276.

Steiner, M., & Neumann, M. (1978). Traumatic neurosis and social support in the Yom Kippur War returnees. Military Medicine, December, 866-868.

USAF Academy Cadet Stress-Induced Immune Suppression

Richard T. Meehan, M.D.
David Jenkins, PhD.
The University of Texas Medical Branch at Galveston, Galveston Texas
Lt Col Robert C. Ginnett, PhD.
Maj Jeffrey S. Austin, PhD.
United States Air Force Academy
Colorado Springs, Colorado CO 80840

Abstract

This multicollaborator research studies the effects of psychological stress experienced by 100 first year male Cadets at the U.S. Air Force Academy on the immune system. The association between susceptibility to EBV infection and individual subjects' perceived stress would be determined and the correlations between blunted in vitro T cell function would be identified. Multiple simultaneous in vitro immune assays would be employed to determine the most suppression (PHA and OKT3 MoAb-stimulated thymidine uptake, expression). We would also determine if stress-induced in vitro immune suppression could be reversed by adding exogenous IL2. The association between circulating ACTH and Endorphin levels and receptor expression on circulating mononclear cells would also be determined to investigate neuropeptide-mediated immune suppression and in vivo peptide hormone receptor expression.

Numerous review articles have summarized evidence from multiple studies that stress can impair in vitro immune responsiveness in humans (Ader, 1981; Colabrese, Kling, & Gold, 1987). The most significant stressors have inclued; bereavement (Bartrop, et al.,1983)), space flight (Taylor & Dardano, 1983), sleep deprivation, academic stress, and hypoxia. The most consistent finding is that cellular immunity as measured by mitogen-induced T cell proliferation is blunted whereas B cell function following these stressors is relatively unimpaired.

The mechanisms responsible for impaired in vitro mitogenic responses following stress, however, have not been identified. Elevated glucocorticoids are probably involved since the administration of hydrocortisone in a dose equivalent to maximal physiologic levels impairs mitogen-stimulated proliferation (Calvano et al., 1987). The in vitro production of several important immunoregulatory cytokines including IL1, IL2, and Gamma interferon is also reduced by adding exogenous glucocorticoids (Synder & Unanue, 1982; Gillis, Crabtree, & Smith, 1979). The exogenous addition of IL2 can also reverse in vitro glucocorticoid-inhibition of T cell proliferation and blunted mixed lymphocyte reactions following thermal injury. Despite the experimental evidence that glucocorticoids induce immune suppression at the level of IL2 production (Munck, Guyre, & Holbrook, 1984), the well known effects of elevated in vivo glucocorticoids including lymphopenia and granulocytosis have been observed only in some studies with stress-induced impaired blastogenesis (Calabrese et al., 1982).

Furthermore, even adrenalectomized animals can exhibit blunted T cell responses following stress (Keller et al., 1983). Therefore, the contribution of hormones other than glucocorticoids mediating stress-induced immune suppression needs additional investigation. A variety of neuropeptide hormones including ACTH and Endorpnins are also elevated during stress and exhibit in vitro immunomodulating activity at physiologic concentrations (Blalock, Bost & Smith, 1985).

Another unanswered question from prior published studies is whether altered in vitro immune responses lead to an increased risk of infectious illness. Other recent studies have reported worse clinical outcomes from infectious complications among trauma patients who exhibited lower in vitro T cell responsiveness (Faist et al., 1986). Increased mortality has also been reported among burn patients who demonstrate the lowest in vitro T cell responses. Animal studies have confirmed that stress can reduce in vivo immunocopetence since animals can be innoculated with pathogens and mortality or infectious disease outcomes can easily be determined (Meehan, 1987). Unfortunately no prospective human stress studies have correlated in vitro immune responsiveness with infectious disease susceptibility.

Our proposal will address the above question based upon an expected incidence of EBV serology conversion (negative to positive) among Air Force Academy Cadets of approximately 20% per year. 209 confirmed cases of mononucleosis were reported between November 1986 and September 1987 among 4,400 Cadets (Drs. Sherry and Schaad - USAFA personal communication). Approximately 893 AFA cadets should have been seronegative (at risk) based upon EBV serology incidence data reported from 1,400 West Point Cadets who were followed 4 years (Kasl, Evans & Neiderman, 1979).

Design

PHASE I August - September 1988

The 100 cadet volunteers would complete monthly questionnaires x 11 using their personal computers. They would also have 10 mls of blood drawn (pre-stress) before 9 a.m. during the first week of orientation (20 subjects/day on 5 consecutive days) in cooperation with the Cadet clinic. The phlebotomy would be repeated 4 weeks later during the end of Basic Cadet Training in Jacks Valley (stress). The PHA-stimulated thymidine uptake assays would be performed in the Biology Labs with slides analyzed at NASA/JSC by digital image analysis.

PHASE II March - June 1989

A cohort of 20 cadets and 10 control subjects would have blood drawn on 4 occasions to coincide with 2 most stressful periods (Pre-recognition and following Survival Training) and 2 less stressful control periods (pre-Spring blood and during graduation exercises). All 30 subjects would have 30 mls of blood drawn on the same day (before 9 a.m.) to minimize assay variability and time in the cadet clinic and hospital phlebotomy areas. Since usually 7-3% of 4th class cadets will drop out before completing BCT, we will defer selecting the 20 cadets for these more complex immune assays (technically difficult and labor intensive) until March. We will select the 10 cadets who demonstrated the greatest and the 10 with the least changes in PHA-stimulated thymidine uptake during BCT to comprise a group of highly stress sensitive (hi-sensitive) and insensitive individuals (lo-resistant) for subsequent study. At the study conclusion an additional 5 mls will be drawn from the remaining 80 cadets so that EBV serology data will be available on all 100 subjects.

113

References

Ader, R. (ed) Psychoneuroimmunology. 1981, New York, Academic Press.

Bartrop, R. W., Luckhurst, E., Lazarus, L., et al: 1977. Depressed lymphocyte function after bereavement. Lancet 1:834-836.

Blalock, J. E., Bost, K., and Smith, E. N. 1985. Neuroendocrine peptide hormones and their receptors in immune system, production, processing and action. J. Neuroimmunology. 10:31-36.

Calabrese, J. R., Kling, M. A. and P. W. Gold. 1987. Alternations in immunocompetence during stress, bereavement, and depression: focus on Neuroendocrine Regulation. Am. J. Psychiatry. 144:1123-1134.

Calvano, S. E., Ablert, J. D., Legaspi, A., Organ, B. C. Tracey, K. J., Lowry, S. F., Shires, G. T., and Antonacci, A. C. 1987. Comparison of numerical and phenotypic leukocyte changes during constant hydrocortisone infusion in normal humans with those in thermally injured patients. Surg. Gynecol & Obstet. 164:509-520.

Faist, E., Kupper, T. S., Baker, C. C. et al. 1986. Depression of cellular immunity after major injury. Arch Surg. 121:1000-1005.

Gillis, S., Crabtree, G. R., and Smith, K. A. 1979. Glucocorticoid-induced inhibition of T cell growth factor production, H: the effect on vitro generation of cytolytic T cells. J. Immunol. 123-1632-1638.

Keller, S. E., Weiss, J. M., Schleiter, S. J. et al. 1983. Stress-induced suppression of immunity in adrenalectomized rats. Science. 221:1301-1304.

Meehan, R. T. 1987. Immune suppression at high altitude. Ann. Of. Emerg. Med. 16:974-979.

Munch, A., Guyre, P. M., and N. J. Holbrook. 1984. Physiological functions in stress and their relation to pharmacologic actions. Endocrine Rev. 5:25-44.

Snyder, D. S., and Unanue, E. R. 1982. Corticosteriods inhibit murine macrophage la expression and interleukin 1 production. J. Immunology 129:1803-1805.

Taylor, G. R. and Dardano, J. R., 1983. Human cellular immune responsiveness following space flight. Aviat. Space Environ. Med. 54:S55.

Stress Inoculation in Operational Settings

Thomas P. O'Hearn, Ph.D.
Front Range Institute
Colorado Springs, Colorado 80919

Abstract

The military has used stress inoculation techniques in a variety of training settings. The results of stress inoculation programs at Lackland AFB and at the Air Force Academy will be reviewed.

Stress Inoculation is not new to the military. Long before the term was used by Meichenbaum (1977) in the early 70s, the military was using stress inoculation methods in varying combat training settings. Using Spielberger's idea (1979) that in order for stress and anxiety to occur, a stressor needs to be perceived as threatening; to the extent that a stressor can be rendered less threatening by cognitive strategies, behavioral techniques, exposure to the stimulus, or a combination of behavioral and cognitive methods then a potential stressor loses some of its power to lessen performance.

Basic training, both in the Air Force and in cadet basic, can be viewed as a stressful event in the lives of both young men and women who have not been exposed to similar training. My work with both basic trainees and Air Force Academy cadets has convinced me that some of the casualties of both programs were the result of the trainee never having experienced this level of stress before and thinking there was "something wrong" with them for feeling what they were feeling. Often when they were assured that what they were feeling was a normal, natural response to a perceived very stressful situation, then their level of stress diminished.

During the 1970s, the Air Force Medical Evaluation Test (AFMET) was initiated at Lackland Air Force Base (O'Hearn and Bloom, 1978). This test program was designed to identify early in the basic training process those individuals who might experience problems in basic training as well as in later Air Force training and work settings. As part of the evaluation procedure, a mental health technician would administer a brief questionnaire on the second day of training. As part of this administration process, the staff at the Mental Hygiene Service felt it would be a good opportunity to briefly discuss with the trainee the fact that basic training would probably contain some stressful moments. Further, the amount of stress perceived had little to do with their worth as persons but may have a lot to do with such things as: prior exposure to similar kinds of stress in sports or military training; prior exposure to communal-type living; the decibel level in the family in which they grew up; and what their levels of expectation were on coming into basic training. During this discussion, the technician briefly pointed out some common stress responses to alert the trainees to their possible stress reactions during basic training.

This procedure was carried out informally for the five years that I was associated with the program at Lackland AFB. The statistics that we kept on the program did not involve stress-related referrals to Mental Hygiene or the General Medical clinics. We did notice a decline in such things as: suicidal-incidents, especially the low lethality ones, and attrition. This is not to say our "stress inoculation" talk was directly responsible for these events as a number of other changes in the basic training process were introduced as a result of the AFMET process.

A second program occurred during basic cadet training at the Air Force Academy during the summer of 1984. Members of the Mental Health clinic at the Air Force Academy were asked to make a 45 minute presentation to the entering class during the first week of training as part of a larger health briefing. Given a sufficient amount of time, the staff presented a mini-stress management lecture to include: the causes of stress particularly as how it relates to basic training; some physiological, psychological, and emotional responses to stress; some cognitive strategies to deal with the stress; imagery techniques; and two relaxation techniques which were done at the close of the lecture. Again, no data were kept comparing attrition, sick call visits, etc. with previous years as research was not the intent of the health program.

Stress inoculation seems to have two main ingredients: the first is the permission to feel stressed in a given situation; the second is the giving of information on successful strategies for dealing with stress. These interventions are preventive in nature and are designed to keep individuals from over-utilizing medical and mental health facilities. Prevention efforts are not foreign to the military, especially in the disease prevention area. Those who have donned the uniform have been given a shot record card which attests to the commitment by the military to inoculate against disease. We have not been as diligent in inoculating personnel against stress. Thinking in terms of prevention rather than remediation or cure might release some creative approaches that are available to all of us to inoculate ourselves, families, and co-workers to the effects of stress.

References

Meichenbaum, D. (1977). Cognitive-behavior modification- an integrative approach. New York: Plenum Press.

O'Hearn, T.P. and Bloom, W. (1978). The Air Force Medical Evaluation Test Program: a three-phase assessment program. In Proceedings of the 25th Annual conference of USAF Behavioral Scientists, Brooks AFB, TX, May, 1978, 105-120.

Spielberger, C. Understanding Stress and Anxiety. London: Harper and Row, 1979.

Who Is At Risk? Health Care Providers Are Potential Casualties Too

A. David Mangelsdorff, Ph.D., M.P.H.
Health Care Studies and Clinical Investigation Activity
Health Services Command
Fort Sam Houston, Texas 78234-6060

Abstract

A number of service professions are stressful because of the nature of the work required. The potential stressors include: event sressors, occupational stressors, and organizational stressors. The results of these stressors can have short term and long term effects.

Health, security, and emergency service occupations are likely to be stressful because of the types of work required. Workloads may shift from low activity with few demands to situations requiring great physical and mental efforts. Workers are highly motivated to get the job done and may disregard their own personal welfare. A number of elements must be considered in service occupations. This paper will review some of the reactions of health service workers in disasters and emergencies; and why professionals and program managers need to be concerned.

Until the 1970s, disaster researchers generally were not trained in mental health methods (Frederick, 1977). Systematic efforts to provide organized mental health services after disasters began in the 1970s, particularly with Section 413 of the Federal Disaster Law of 1974 (Baisden and Quarantelli, 1981). Since then, some considerations for planning, implementing interventions and following survivors/victims for post traumatic effects have been implemented (Mangelsdorff, 1985; Mangelsdorff, King, and O'Brien, 1986).

A National Institute of Occupational Safety and Health survey of a range of occupations showed that of those with the highest incidence of stress related disorders, seven were health care occupations (Schwartz, 1978). A Department of Labor study reported that accident rates both on and off the job were 58% higher for health care personnel than those employed in other service organizations (Calhoun, 1980). In community and psychiatric hospitals, staff stress adversely affected patient care (Calhoun, 1980). Work conditions can have adverse effects on employee productivity and morale (Cooper and Payne, 1978).

Hartsough and Myers (1985) describe selected factors which contribute to putting the service provider at risk as: event stressors, occupational stressors, and organizational stressors. Hartsough and Myers (1985) document event stressors which may affect service providers: personal loss or injury, traumatic stimuli, and mission failure or human error; any or all of these stressors may increase the provider's risk to negative reactions. Reactions to traumatic situations are personal; each individual perceives a situation as a function of his (her) own experiences, training, and feelings.

Emergency service workers are at risk because of the hazardous nature of their occupation. The occupational stressors may include: the working environment, time pressures, and the work load (Hartsough and Myers, 1985). Some of what may result will be short term in duration and other factors will manifest long term effects.

Short Term Responses.

Killian (1952) reported traumatic aftermath in police officers, firefighters, and public utility workers following the Texas city oil port explosion. The rescue workers reported dilemmas between staying on the job or seeing to the welfare of family and friends. Organizations responsible for the supervision of rescue workers must address the potential for disruption of support services during disasters situations caused by trauma in their workers (Diskin, Goldstein, and Grencik, 1977). Workers at emergencies sometimes feel compelled to deny their feelings and inadequacies; their perceived role conflict does not allow them to show weaknesses. Emergency workers are trained to suppress emotional reactions during and for a brief time after an incident. Natural feelings of denial and avoidance are predominant during the first 24 hours. The workers may attempt to intellectualize the incident. As their cognitive processing decreases, intense feelings may then come to the surface. The rapid onset of shock and numbed feelings may prevent the emergency worker from effectively performing their work during an emergency.

Long Term Effects.

Health care professionals have long ignored the needs of persons who act as rescue workers during crisis situations despite later occurrence of medical and mental disorders (Raphael, 1980). The National Institute of Occupational Safety and Health (McLean et al, 1978) documented that certain rescue occupations involve high stress and make the workers at risk for diseases of adaptation. Research has found that police officers experience stress-precipitated diseases at rates higher than the general population, have higher rates of heart disease and stomach disorders, higher divorce rates, higher suicide rates, and significant health, substance abuse, and marital problems (Kroes and Hurrell, 1975). The police and other emergency service providers do not become immune to the stresses of their occupation. Given the cumulative nature of stress, the onset of a sudden traumatic event could trigger a stress reaction.

Burnout.

Burn-out refers to changes over time in human service providers toward their work and their clients. The negative behaviors include: detachment, cynicism, lack of empathy, boredom, pessimism, fatigue, negativism, blaming clients. The burn-out may be a response to long term stressors, organizational unresponsiveness, or unrealistic expectations (Winget and Umbenhauer, 1982).

Why the Concern?

Turnover of highly trained, skilled personnel is expensive in terms of organizational and personnel costs. The amount of time required to train new personnel and the additional burden placed on experienced workers while the new ones are being trained, can create dissatisfaction and additional personnel turnover. While there has been considerable research on turnover (Porter and Steers, 1973, 1977; Locke, 1976; Mangelsdorff, 1978, 1984, 1985), and on occupational stress (Cassel, 1976; Cobb, 1976; LaRocco, House, and French, 1980), there has been limited work to separate out occupational stressors from non-work stressors which would allow program managers and practitioners to reduce job-related turnover.

Program managers and supervisors need to be concerned with organizational stressors. The most obvious problems concern role conflicts and role ambiguity. Role conflicts can include: perceived differences in professional versus personal behaviors, divided loyalty to client or to the organization, work versus home and family responsibilities, worker and supervisor, or among coworkers. Managers need to be supportive of their workers and provide an environment that minimizes potential role conflicts and stressors.

References are available on request from author

Enhancing the Selection of Air Force Personnel: New Research Directions

Thomas W. Watson
Linda R. Elliott
Air Force Human Resources Laboratory

Victor H. Appel
The University of Texas at Austin

Abstract

This paper describes accomplishments to date and anticipated future directions of a new research program at the Air Force Human Resources Laboratory (AFHRL). The aim is to develop promising new predictors of success of Air Force officers (both commissioned and non-commissioned) to enhance selection. Particular interest is focused on selecting candidates who show promise for becoming good managers/ leaders, and who are likely to be strongly committed to the Air Force as an institution. The research described is shifting from a conceptual to an instrumentation phase in which the scope of activities is being delimited.

Two instruments form the primary basis for Air Force personnel selection: the Air Force Officer Qualifying Test (AFOQT) for officers and the Armed Services Vocational Aptitude Battery (ASVAB) for enlisted personnel. The ASVAB is also used in the other services of the Department of Defense. As an aid in personnel selection, these tests effectively assess general learning or academic ability using paper-and-pencil subtests combined to form math and verbal composites. Other composites are used for classification, but the math/verbal composites allow the Air Force to identify personnel for selection who are sufficiently gifted to benefit from training and to perform technical aspects of their jobs well, if they are motivated to do so.

Although these instruments are useful selection devices, they do not attempt to tap certain attributes or abilities which may contribute to successful performance as an Air Force member. For instance, they do not assess the likelihood that a candidate will have: (a) the leadership or management qualities requisite to be a good officer, such as organizational, interpersonal, or communication skills, or the (b) commitment motivating performance toward the attainment of shared individual and organizational goals. It was our desire to enhance selection by identifying or developing such measures to complement those currently used.

Conclusions Based on Our Exploration of the Literature

Since it is wise to have a good conceptual foundation before developing measures, the summer of 1987 was spent examining literature in the following areas relevant to our concerns: leadership/management and motivation/ commitment. Space limitations preclude a complete discussion of this literature here. Interested readers should consult Elliott & Watson (1987) Carson & Sperl, (1987), and Watson (1987) for more detail.

Leadership and Management Literature

Paper-and-pencil measures. In exploring the literature we found that little was new until recently (Watson, 1987) and that few good paper-and-pencil instruments exist to measure leadership or management propensity at the point of selection. Possible exceptions are the Tacit Knowledge Inventory for Managers (TKIFM) published by the Psychological Corporation of San Antonio, Texas (Wagner & Sternberg, 1985; 1987) and selected new subtests under development for AFHRL by Psychometrics, Inc. of Sherman Oaks, California. Carson & Sperl (1987), and Eddy (1987) describe the utility of the TKIFM, while Steuck (1987) describes the usefulness of the latter instruments.

As we contemplated the development of new paper-and-pencil instruments, we recognized it will be a challenge to develop measures resistant to being "gamed" by recruits intent on gaining admission into the Air Force. We decided that an indirect measure, such as a biodata form, would be a promising initial approach to pursue in tapping leadership/management propensity and commitment prior to organizational entry.

Behavioral measures. The assessment center literature provided a source of behavioral assessment techniques. This technology showed great promise when applied in selected situations. Despite its high cost, it can be cost effective (Elliott & Watson, 1987). It may be practical to explore the merits of elements from this technology for use in selection of Air Force members. The in-basket exercise, the panel interview, or selected communications exercises are especially promising. However, exploration of a complete assessment center approach would be costly and might not gain support.

Commitment/Motivation Literature

Our interest in commitment (which implies work motivation and the consonance of values between individuals and their organizations) found support in the literature and was related to leadership/management since commitment seems to be closely tied to a new conceptualization of leadership called transformational leadership (Bass, 1985; Watson, 1987). Thus, at the point of organizational entry, it may be worthwhile to measure the consonance of values between recruits and the Air Force, and the probability of recruits becoming committed to the Air Force in an organizational or moral sense (Etzioni, 1961; Mowday, Porter & Steers, 1982). The focus on commitment is in keeping with Moskos' (1986) emphasis on the desirability of a stronger institutional than occupational orientation among military personnel. It is also consonant with Watson's (1986) finding that commitment is the most critical factor influencing the retention of Air Force enlisted personnel.

Current Research Activities

Promising Cognitive Measures of Leadership and Management

The following five subtests under development by Psychometrics, Inc. seem promising for selecting personnel with leadership/management potential.

Text Editing: Examinees are asked to choose the most correct relacement sentence for a flawed sentence in a paragraph. This test appears to be relevant to written communication ability critical for managers and leaders.

Management Organization: A management problem is described briefly and the examinee must choose an option describing the best-ordered activities to be followed in addressing the problem. This subtest strives to tap the important managerial planning activity of prioritizing tasks.

Deductive Reasoning: This tutorial subtest contains a variety of Venn-diagram problems from set theory. It is believed relevant to managers since they must be sufficiently deductive to resist acceptance of erroneous conclusions.

Flowchart Reading: This subtest contains problem solving situations presented in a format similar to the flowcharts used by computer programers. The examinee must select from among alternative paths at incomplete decision points. This subtest appears to tap a manager's ability to make rapid, effective decisions.

Deciphering Languages: Examinees are presented a nonsense syllable "language" and must decipher what English-language words they stand for, using an implicit decision rule. As with Deductive Reasoning, this subtest attempts to measure reasoning ability, an important managerial attribute.

Administering the TKIFM to Air Force Personnel

Tacit Knowledge is presumed to be a component of practical intelligence (not tapped in academically oriented intelligence tests) which is acquired and used but not openly expressed or stated. Thus it is "picked up" without being explicitly taught (Wagner & Sternberg, 1985). Eddy (1987) is administering the Tacit Knowledge Inventory for Managers (TKIFM) to Basic Military Training students who have also taken the ASVAB. Her initial purpose is to test whether the TKIFM is indeed tapping a construct not currently measured by the more traditional ASVAB. For instance, it is possible, contrary to the intent, that the TKIFM is tapping a verbal component. This component may be similar to that measured by the ASVAB, rather than tacit knowledge. If, however, the two tests are measuring different abilities, then further research would determine the merit of using the TKIFM as an adjunct to the ASVAB or AFOQT for enhanced selection.

Proposed Research Activities

Development of a Biodata Measure of Leadership and Management Propensity

Problems, such as racial or ethnic bias and the focus on life experiences not under personal control, have plagued the use of biodata inventories in the past (Stricker, 1987; Laurence & Means, 1987). However, we believe that many of these problems could be overcome through careful development of such an inventory. Also, advantages are clear. Experiences related to leadership/management can be measured indirectly and in a potentially verifiable manner. Thus, it is reasonable to assume the number of prior leadership roles assumed, whatever their context, should be predictive of degree of leadership competence. Prototype biodata forms will be developed, pretested, refined, and then administered to approximately 2000 personnel in each of the following criterion groups: Above- and Below-the-Zone promotees to Major and comparable groups of promotees to Chief Master Sergeant. Item response data will be analyzed to identify items which discriminate between the criterion NCO and

officer groups and which have no adverse impact. Promising items will form instruments and item pools for further research and possible operational use.

Development of Measures of Commitment and Values

Organizational commitment is typically defined in tripartite terms (Mowday, Porter & Steers, 1982), having the following components: (a) values (consonance between individual and organizational values or goals), (b) motivation (willingness to exert effort toward organizational goals) and (c) retention (desire to maintain membership). Although commitment can be maintained and enhanced (or eroded) on the basis of organizational experiences, Mowday, Porter and Steers provide evidence for anticipatory (pre-employment and job choice) influences on commitment. Thus, candidates have differential propensities for becoming committed to organizations before they enter, and are already committed to greater or lesser degrees upon entry. This differential propensity for commitment is probably tied to values which impact personal and social norms. However, unlike norms -- which are tied to actors in specific social contexts -- "values are standards of desirability that are more nearly independent of specific situations" (Schwartz, 1977, p. 232). They are, therefore, probably stable, and brought by recruits into the organization. If a match exists between individual and organizational values, commitment components are likely to be strong, and resistant to erosion.

Measuring commitment and values at the point of entry will pose a difficult challenge. Commitment or values are relatively easily measured after organizational entry when respondents don't feel as compelled to respond in a socially desirable way. However, recruits may try to appear committed when they are not, or to express shared values not actually held. One way to avoid this potential problem would be to tap commitment and value consonance through biodata items, just like leadership or management propensity.

Development of Additional Measures of Problem Solving and Written Communication

A recent survey (Fenno, 1985) supports the need for communication and interpersonal skills across Air Force officer occupations and grades, and helps to specify the specific competencies involved. We plan to explore ways of assessing communication, interpersonal skills, and problem solving using a paper-and-pencil format different from those proposed by Psychometrics, Inc. Of most interest is a simplified in-basket test, based on tasks and situations encountered by junior commissioned and non-commissioned officers. Multiple choice items will be developed to measure comprehension, interpersonal sensitivity, written communication, planning, and prioritizing skills. Content validity of the items will be based on job analysis and survey feedback data from officers and NCOs. Concurrent and predictive validity data will be collected in the form of correlations between test performance and multiple criteria such as operational enlisted/officer performance ratings, experimental performance ratings, and other factors described by Scott (1984) as most influential to an Officer Promotion Nomination Board.

<div align="center">References</div>

Bass, B. M. (1985). Leadership and performance beyond expectation. New York: Free Press.

<div align="center">123</div>

Carson, A. D. & Sperl, T. C. (1987, November). The use of noncognitive paper and pencil measures of leadership potential. In V. H. Appel (Chair), New directions in leadership and managerial selection, Symposium presented at the annual convention of the Texas Psychological Association, San Antonio, TX.

Eddy, A. S. (1987). The relationship between the Tacit Knowledge Inventory for Managers and the Armed Services Vocational Aptitude Battery. Unpublished master's thesis in progress. San Antonio: St. Mary's University.

Elliott, L. R., & Watson, T. W. (1987, November). How useful can we expect assessment center derived measures of leadership and managerial ability to be? In V. H. Appel (Chair), New directions in leadership and managerial selection, Symposium presented at the annual convention of the Texas Psychological Association, San Antonio, TX.

Etzioni, A. (1961). A comparative analysis of complex organizations. New York: Free Press.

Fenno, C. R. (1985, March). A profile of the communications tasks of Air Force officers. (AU-AFIT-LS-1-85, AD-A163 476). Wright Patterson AFB, OH: Air University.

Laurence, J. H., & Means, B. (1985, July). A description and comparison of biographical inventories for military selection (FR-PRD-85-5, AD-A157 364). Human Resources Research Organization, Alexandria, VA.

Moskos, C. C. (1986, July). Institutional and occupational trends in military organization. Unpublished Technical Report. Bolling AFB, MD: Air Force Office of Scientific Research.

Mowday, R. T., Porter, L. W., & Steers, R. M. (1982). Employee-organizational linkages: The psychology of commitment, absenteeism, and turnover. New York: Academic Press.

Schwartz, S. H. (1977). Normative influences on altruism, In L. Berkowitz, (Ed.), Advances in experimental social psychology (V 10). New York: Academic Press.

Scott, L. M. (1984). Identifying variables influential to an officer promotion nomination board (AD-P003 33410). Randolph Air Force Base, TX: Air Force Manpower and Personnel Center.

Steuck, K. W. (1987, November). Air Force Officer Qualifying Test (AFOQT): Promising cognitive measures of leadership and management In V. H. Appel (Chair), New directions in leadership and managerial selection, Symposium presented at the annual convention of the Texas Psychological Association, San Antonio, TX.

Stricker, L. J. (1987, October). Developing a biographical measure to assess leadership potential. Paper presented at the annual meeting of the Military Testing Association, Ottawa, Canada

Wagner, R. K & Sternberg, R. J. (1985). Practical intelligence in real-world pursuits: The role of tacit knowledge. Journal of Personality and Social Psychology, 49, 436-458

Wagner, R. K & Sternberg, R. J. (1987). Tacit knowledge inventory for managers. San Antonio: The Psychological Corporation

Watson, T. W. (1986, August). Full-scale test of an empirical model of turnover. Paper presented at the 1986 national meeting of the Academy of Management, Chicago, IL.

Watson, T. W. (1987, November). Leadership versus management: Conceptual similarities and differences. In V. H. Appel (Chair), New directions in leadership and managerial selection, Symposium presented at the annual convention of the Texas Psychological Association, San Antonio, TX.

124

Work Dimensionality Among United States Air Force
Enlisted Specialties

Maj William E. Wimpee
United States Air Force
North Carolina State University

J. W. Cunningham, Ph.D.
North Carolina State University

Lt Col Rodger D. Ballentine
Air Force Human Resources Laboratory

Abstract

The purpose of the reported research was to derive a
set of general work-activity factors underlying United
States Air Force enlisted specialties. The resultant
factors and a specialty cluster structure based on the
factors should provide a useful prototype for the
development of a quantitative occupational taxonomy.
Among the potential areas of application for such a
taxonomic system are career exploration and guidance, job
placement and transfer, general vocational capability
training, and job evaluation.

Historically, job analysis in the U. S. Air Force has been based
on the task inventory questionnaire method (Christal, 1974). As
developed by the Air Force, the task inventory approach has proven
to be a powerful tool in training program development and job
classification, and has been widely adopted in government, industry,
and occupational education. The specificity of task descriptions,
however, limits their usefulness in comparing jobs across a broad
occupational spectrum. The research reported here was designed to
test the applicability of a more general job analytic approach in
developing a quantitative taxonomic system for Air Force enlisted
specialties.

Method

Instrument

The primary research instrument was a structured job analysis
questionnaire, the General Work Inventory (GWI), developed by the
second and third authors (Cunningham and Ballentine, 1982). The GWI
derives from and is intended to replace a longer research
questionnaire, the Occupation Analysis Inventory (OAI; Cunningham,
Boese, Neeb, and Pass, 1983). The GWI is considerably shorter than
the 617-item OAI and is written at a less technical level. Its 266
job-rating items, or work elements, are based on selected factors
and modified elements from the OAI, in addition to information from
other sources. The GWI is organized into eight sections of work

125

elements on which jobs are rated: (A) Sensory Requirements,
(B) Information Elements, (C) General Mental Requirements, (D)
General Physical Requirements, (E) Physical Work Activities, (F)
Interpersonal Work Activities, (G) Work Conditions, and (H) Job
Benefits/Opportunities.

The GWI was designed to achieve as much specificity in
description as possible, while retaining applicability to the
general population of jobs; that is, to approach the limit in
descriptive specificity separating generally applicable devices
(such as the Position Analysis Questionnaire; McCormick, Jeanneret,
and Mecham, 1972) from devices designed for restricted job
categories (such as the Air Force job inventory questionnaires). In
contrast to the Position Analysis Questionnaire, which was designed
to describe the general "worker-oriented" characteristics of jobs
(primarily for purposes of job-component validity and job
evaluation), the GWI represents an attempt to capture some of the
more specific technological (or "job-oriented") content of jobs, in
addition to some of the basic human requirements.

Two rating scales are used in the GWI. The elements in GWI
sections A thru F (the information and activities variables) are
rated on the Part-of-the-Job scale. The elements in sections G and
H (the work conditions and benefits/opportunities variables) are
rated on the Extent-of-Occurrence scale. The Part-of-the-Job scale
asks the rater to consider three factors in determining an element's
part in the job: (1) how important or critical the element is to
the job, (2) how often the element occurs, and (3) how much time
the job holder spends with the element. This scale is a
modification of Hemphill's (1959) Significance scale, which despite
its complexity, has been found by previous investigators to have
substantial reliability (Cragan and McCormick, 1967; Tornow and
Pinto, 1976). The scale has nine possible responses, ranging from
"1" (an extremely small part of the job) to "9" (an extremely large
part of the job).

The Extent-of-Occurrence scale was designed to apply to work
conditions and job benefits/opportunity elements for which the
Part-of-the-Job scale seemed inappropriate. This scale also asks
the rater to consider three factors: (1) how often the element
occurs, (2) how much time the element occurs, and (3) the level at
or degree to which the element occurs. This scale has nine possible
responses, ranging from "1" (an extremely small extent of
occurrence) to "9" (an extremely large extent of occurrence).

Sample

The GWI and accompanying position rating instructions were mailed
to approximately 2,500 Air Force enlisted personnel employed in 175
specialties. A total of 2,141 respondents completed the rating
task. In addition, a follow-up sample of 112 respondents rerated
their positions two to three months after completing their initial

126

GWI ratings.

Analyses and Results

Based on the follow-up sample data, a rate-rerate correlation was computed for each GWI element and for the GWI element rating profile for each respondent. The mean rate-rerate correlations for the 266 GWI elements was .62, and the mean element profile correlation for individual respondents was .74.

Seven sections of the GWI elements (Sections B thru H) were factor analyzed based on 2,141 position ratings by enlisted personnel in 175 specialties. Section A, Sensory Activities, was excluded from the factor analyses under the assumption that the basic sensory capacities are unidimensional. A separate factor analysis was performed on the elements from each of the seven GWI sections, and an overall analysis was performed on the elements from all seven sections combined. In addition, a general higher-order factor analysis was performed on the resultant sectional factors using both estimated and exact factor scores of respondents. The analytical method in all cases consisted of principal components analysis followed by varimax rotation. The number of factors rotated in each analysis was determined by an examination of the eigenvalue plot. In most cases, two or three different numbers of factors were rotated, and the most meaningful solution was retained. All of the factor analyses were carried out separately with the total sample of respondents and with each of two comparable subsamples of 1,045 respondents. Analyses of factorial replication, or stability, were performed involving the computation of Tucker's congruence coefficients (Gorsuch, 1974) for pairwise matchings of rotated factors between the two subsamples.

The resultant factor structures were, for the most part, meaningful and replicable. Altogether, 59 interpretable factors emerged from the seven separate sectional component analyses. Fifty-five of those factors had congruence coefficients $\geq$ .70. Moreover, the jobs with the highest factor scores seemed to logically characterize the factors. The higher-order analysis of the estimated scores of 55 interpretable sectional factors with relatively high congruence coefficients ($\geq$ .70) resulted in a solution of 15 interpretable factors, 14 of which had congruence coefficients $\geq$ .85 and accounted for 66.57 percent of the total variance. The overall factor analysis, involving seven GWI sections combined, produced a 44-factor solution. Thirty-six of those factors, accounting for 56.40 percent of the total variance, were interpretable.

Discussion

The GWI factors will be used as a basis for clustering Air Force enlisted specialties. In an initial methodological study involving 40 specialties, the second and third authors will compare separate results using the sectional factors, overall factors, and GWI items as profile variables. They will also compare three profile

127

association measures: squared Euclidean distance, product-moment correlation, and percent overlap. The criterion in this study will be a 40-by-40 matrix containing mean pairwise similarity ratings of the specialties by job experts. The combination of a set of profile variables and an association measure that most closely approximates the criterion matrix will be applied in a subsequent large-scale cluster analysis of 175 specialties. The large-scale cluster structure will be tested for replicability by comparing the results obtained from two comparable subsamples of respondents. The replication analyses will involve (1) correlations of cluster centroids between the two subsamples and (2) cross-classification contingency analyses to determine the extent of joint specialty membership between matched subsample clusters. In addition, the cluster structure will be tested for agreement with its association matrix (based on cophenetic correlation).

It is hoped that the GWI factors and the enlisted specialty cluster structure derived from them will serve as a useful prototype for the development of a quantitative occupational taxonomy with potential applications in career exploration and guidance, job placement and transfer, and possibly the development of general vocational capability training curricula based on the clusters and their salient GWI factors. Ultimately, we hope to apply McCormick's (1979) job component approach in linking the GWI factors to personal attributes, such as interest scales in the Air Force's VOICE (Vocational Interest for Career Enhancement) and the tests in the ASVAB (Armed Services Vocational Aptitude Battery). A work classification system consisting of broadly applicable GWI work descriptors related to key personal attributes and occupation-specific task data would be a very powerful manpower, personnel and training management tool.

References

Christal, R. E. (1974). The United States Air Force occupational research project (AFHRL-TR-73-75), Air Force Human Resources Laboratory, Occupational Research Division, Brooks Air Force Base, TX. (NTIS No. AD-774 57488GI).

Cragan, J. R., & McCormick, E. J. (1967). Job inventory information: Task reliabilities and scale interrelationships (PRL-TR-67-15). Lackland Air Force Base, TX: Personnel Research Laboratory. (NTIS No. AD 681 509).

Cunningham, J. W., & Ballentine, R. D. (1982). The general work inventory. Raleigh, North Carolina: the authors.

Cunningham, J. W., Boese, R. R., Neeb, R. W., & Pass, J. J. (1983). Systematically derived work dimensions: factor analysis of the Occupation Analysis Inventory. Journal of Applied Psychology, 68, 232-252.

Gorsuch, R. L. (1974). Factor Analysis. Philadelphia: W.Saunders.

Hemphill, J. D. (1959). Job descriptions for executives. _Harvard Business Review, 37,_ 55-67.

McCormick, E. J. (1979). _Job analysis: Methods and applications._ New York; AMACOM.

McCormick, E. J., Jeanneret, P. R., & Mecham, R. C. (1972). A study of job characteristics and job dimensions as based on the Position Analysis Questionnaire (PAQ). _Journal of Applied Psychology, 56,_ 347-368. (Monograph).

Tornow, W. W., & Pinto, P. R. (1976). The development of a managerial job taxonomy: A system for describing, classifying, and evaluating executive office positions. _Journal of Applied Psychology, 61,_ 410-418.

Psychology in the FMOD - Officer Career Planning
in the German Federal Armed Forces for the 90s

Dr. Friedrich W. Steege
Federal Ministry of Defense - P II 4
Bonn, Federal Republic of Germany

Abstract

The Military Psychology Section of the Federal
Ministry of Defence (FMOD) has been tasked to in-
vestigate current problems concerning the number
and the quality of commissioned officers in the
German Federal Armed Forces (GFAF) in the years to
come. This paper is providing some information
about the background and the planning phase of a
survey conducted in our armed forces in December
1987 / January 1988. The aim of the survey is to
get additional evidence with a view to decisions to
be taken to secure a sufficient number of tempora-
ry-career officers for a commission, especially in
technical specialties, in the medium term.

In the early seventies, two Armed Forces universities
came into being. Among other things, they serve the purpose of
making an officer career in our armed forces more attractive
and of increasing the quality of the officers corps. Since
then, the number of applications and the selection ratio has
been very satisfactory. Furthermore, a high level of academic
and professional standards has been achieved. Nevertheless,
there are still several problems concerning the number of fu-
ture commissioned officers.

Assignment Figures

The GFAF need some 650 commissioned officer candidates
per age group for line duty in order to maintain operational
readiness and achieve a balanced age structure. Only one third
of them are enrolled from the outset as commissioned officer
candidates. Most officers must qualify for commissioned status
by their performance in the armed forces, serving first as tem-
porary-career personnel.

In 1985 and 1986, some 60 % of the applicants and some 70 % of the accepted officers were graduates from the *Bundeswehr* universities. About 27 % *of the officers* transferred to commissioned status were graduates from technical engineering and computer science faculties. About 40 % of the applicants did not enrol in, or complete, courses of study.

Problem

The assignment of commissioned officers to special assignment categories is no general problem yet. We are, however, currently facing discrepancies caused by the requirements of our personnel management and the special aspirations of officers having graduated from the Bundeswehr universities. This is especially true with respect to technical engineering and computer science courses of study. There are indications that in these faculties the majority of students regard their diplomas as a prerequisite for a job in private industry rather than for a commission as officer. This may cause difficulties in satisfying the *Bundeswehr* requirements for commissioned officers in the medium term.

Reasons and Conflicting Goals

The following <u>reasons</u> causing the problem situation have been determined:

(i) "Workplace" *Bundeswehr*
   - Discrepancy between university courses of study and reality of the job, i.e. jobs differ considerably from contents of study courses.
   - Discrepancy between wage expectation after graduation and income offered by armed forces.
   - Relatively low job satisfaction of graduated officers.

(ii) Personnel Management
   - Negative effects emanating from the alleged seniority principle.
   - Negative repercussions of the assignment backlog.
   - Requirements of mobility.

(iii) Societal Factors
   - Family situation.
   - Excessive duty hours as dictated by the mission.
   - Little attractiveness of many garrisons.

These reasons are basically determined by at least three conflicting goal areas:

(i) Personnel management aspects vs. self-comprehension of the armed forces universities.
The personnel management of the armed forces is directed by the necessities of military assignments. The self-comprehension of the armed forces universities is primarily shaped by the comparability with scientific education at civilian universities. I.e., they view first of all their genuine (academic) role.

(ii) Effectiveness vs. attractiveness of an officer career in the GFAF.
Positions that require graduation are expected to guarantee higher prestige and better advancement possibility. If a temporary-career officer feels misled in these expectations, he may tend to leave the armed forces and not seek a commission. This may be particularly true if private industry is trying to offer contracts long before the end of an officer's temporary term.

(iii) Military ethics vs. integration into society.
Officer candidates tend to adopt the wrong attitude that courses of study are the main element of their profession during their long years on the university campus. Military leadership has on the other hand to further the attitude that courses of study are part of the officer training. The principal aim of this training is to prepare the officer for his profession, i.e. for his military duty. Therefore, military leadership has to use every possibility to educate the individual officer with respect to the legitimate requirements and the genuine particularities of the officer profession.

Furthermore, the armed forces have to face the well-known difficulties caused by societal developments like decreasing mobility, aspirations of husbands in (civil) occupations (two career couples), and other motivational factors.

132

Proposed Measures

An analysis of official reports submitted to the FMOD has identified several measures which ought to be taken into consideration. These measures are mainly aimed at:

(i) Recruiting
- Win more applicants for commissioned officer status.
- Improve selection methods.

(ii) Advertising, Counseling
- More frequent career counseling and courses of practical work or on-the-job training.
- Advertising; Posts offered on a competitive basis.
- Better information for/by commanding officers.

(iii) Career Regulations
- More flexibility in general.
- Increase in the number of commissioned officers without graduation including officer specialists.
- Earlier assignment of commissioned officer status.

(iv) Budgetary Measures
- Improve bonus system.
- Improve general payment structure.

(v) Identification with professional standards
- Increase professional satisfaction in general.
- Increase acceptance of personnel management.
- Better professional motivation of superiors.

Evaluation

The evaluation has led to the general conclusion that there is too little evidence as to what decision the majority of the respective officers will take. The reasons gathered up to now leave open what weight each motive has regarding the decision by the individual officer. Therefore, a representative survey has been designed to investigate the motives/reasons of the respective officers.

What we expect is to find more information about possible reasons for the decrease in the number of applicants for a commission, especially in the technical faculties. We included items regarding the competition with the private market especially in the field of these technical areas. We have stressed comparisons with respect to career and income opportunities. We will evaluate whether it is sufficient, or a futile hope, to emphasize the genuine particularities of the officer career, taking into account the fact that the officer career may be tantamount

to smaller earnings than in private industry. The appeal to soldierly ethics may to a certain extent only be realistic if an officer's economic situation is at least comparable with other occupational fields.

We also have to find out what repercussions the earlier finding will have, namely that most of the temporary-career officers who left the service in 1984 and 1985 have found an adequate, or even a better, job in the civilian market after leaving the *Bundeswehr*. This, again, holds especially true for graduates of technical courses of study.

Action

In December 1987, the questionnaire was mailed to about 4,200 officers or officer candidates (students at the armed forces Universities) of those five age groups that have commenced studies in 1981, 1982, 1983, 1985, and 1987). We included officers / officer candidates (i) after having entered an AF university, (ii) after having received the first university degree, (iii) after graduation, and (iv) after one to two years following their return to military life. We are just receiving the completed questionnaires and doing the first inspective analyses. Results and interpretations will not be available until April this year.

Factors Influencing the Prediction of
Target Detection Performance[1]

Lieutenant Colonel Timothy R. O'Neill, Ph.D.
Department of Behavioral Sciences and Leadership
United States Military Academy

Abstract

The soldier's ability to detect a target visually at extended ranges becomes critical as the lethal range of weapons increases. In this study, we identify the critical factors that influence battlefield target detection, with the objective of making better use of manpower by selecting soldiers with high aptitude. Evaluation of a wide range of measures indicates that the principal faculties that determine detection under such conditions are spatial acuity, measured as contrast sensitivity in low, mid- and high-frequency ranges, and the size of the visual lobe, or working field of view. Emerging results suggest that a short battery of tests may explain roughly 64 percent of variance in target detection skill.

Direct-fire weapons systems such as the TOW represent a trend in increased range and lethality that already appears to exceed the ability of many gunners to acquire and engage a ground target. Typical military targets at ranges beyond 2000 meters present an optical event with extremely small angular subtense, and countersurveillance efforts degrade their visible signatures even further. Even with the aid of sight optics, a full flank presentation at longer missile ranges will present a full flank presentation of no more than a few degrees.

Predicting a soldier's ability to detect and identify targets under such conditions provides one answer to the problem; but finding such predictive measures has proven somewhat elusive, even if we ignore nonstimulus variables. There is, however, evidence of some factors that may be powerful predictors.

Ginsburg and associates (1982) and O'Neill (1986) report a strong correlation between visual acuity measured as contrast sensitivity for drifting sinewave gratings and ability to detect such targets. Ginsburg's method employed US Air Force pilots in an aircraft simulator, and used as a dependent measure their discrimination distance for aircraft targets or small angular subtense, and resulted in a rank-order correlation of roughly .8 between detection distance and contrast sensitivity. O'Neill used degraded target shapes embedded in a background of visual noise, and noted similar relationships.

A second possible factor is the visual lobe. This is a problematic term, since it seems to imply an anatomical locus on the cortex; in fact, it is a construct representing a "working field of view," centered on the fovea, within which complex pattern detections are possible (Bloomfield, 1975; Engel, 1976; Widdell and Kastner, 1981; Gallwey, 1982; Bellamy and Courtney, 1981). In the last study, which exemplifies the run of results, the authors report a correlation of .92 between visual lobe size and visual search performance.

The presumed relationship was proposed by Widdell and Kastner (1981): since visual lobe size varies, we might expect that observers with larger lobes might be better target detectors based on simple probability of including a critical piece of the optical array in one fixation -- in essence, take a larger visual "gulp" than an observer with a smaller visual lobe.

Our hope was that a combination of these two factors -- contrast sensitivity and visual lobe size -- might yield strong predictability, since each has been shown to correlate strongly with target detection and there is no reason to believe the two faculties are colinear.

Method

Observers

Thirty volunteers from the staff and faculty, United States Military Academy and 1st Battalion, 1st Infantry participated in the study. All Observers were pretested for Snellen acuity, color vision, and stereopsis.

Apparatus

Contrast sensitivity was measured in two ways: by computer-generated CRT display (Innisfree image generator controlled by an IBM PC XT/Cyborg 911 laboratory computer, displayed on a Tektronix 608 display monitor with P31 phosphor, and by the Vistech VCTS measurement system. The computer measure used gratings of 2, 6 and 18 cycles/degree; the VCTS uses 1.8, 3, 6, 12 and 18 cycles/degree.

Visual lobe was measured using a locally-developed test (VLT). This consisted of a computer-driven display with a central fixation point, around which letters flashed briefly at various distances from the center along the horizontal and vertical meridians; the Observer viewed the test on a 19 inch composite monitor at a viewing distance of 1.3 meters.

Target detection performance was measured in three ways, using a visual test system designed at West Point and consisting of an acoustically and electrically sealed suite with rear-projection screen and provision for measuring a variety of perceptual and physiological activities. First, the traditional measure of visual detection response time was tested using rear-projected 35mm slides of targets in natural settings (military vehicles set in tree lines at ranges varying from 800 to 1500 meters); a keypress signalled detection, and score was the mean detection time for 30 presentations.
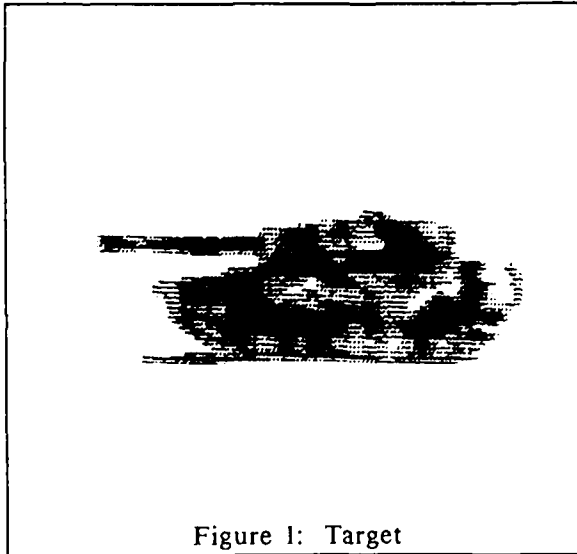
Figure 1: Target
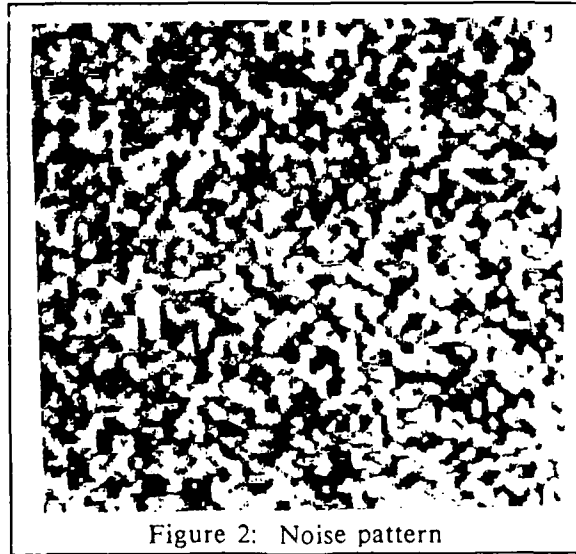
Figure 2: Noise pattern

Detection time, however, does not allow for nonstimulus variables. For this reason, we also used slides described by O'Neill (1986): degraded targets embedded in a noisy background (Figures 1 and 2). Slides appeared for 100 msec, and the Observer reported whether the scene contained a target or did not, and the quadrant in which the target appeared. There were 30 slides, including 8 null targets. This resulted in two-state ROC curves for each Observer, the empirical d' providing a measure of detection ability that included both hits and false alarms.

Finally, we measured the number of fixations and dwell times for 30 slides similar to those in the first target trials, using an ASL 1996 eye view monitor.

Procedure

Observers completed the test in two sessions. In the first, the contrast sensitivity and visual lobe tests were administered, along with the pretests. In the second, the three target detection tests were conducted.

In the second phase, Observers were seated in an adjustable chair at a distance of 2.6 meters from a rear-projection screen. In the first test (eye movements), the pupil calibration required relative head immobility for a series of 10 visual search slides. The Observer viewed the scene, freely searching for the target, then on detection signalled a stop in eye measurement and identified the location of the target using a superimposed coordinate system; this was repeated for ten segments. At this point the eye monitoring ceased and free head movement was restored.

The 30 timed trials then proceeded, with target detection recorded by a hand switch. After each trial, the Observer verified target location using the coordinate system.

Finally, the 30 signal detection exposures were presented; after each the Observer reported "yes" or "no" to the presence of a target and verified location of positive detections using the coordinate system.

Results and Discussion

Emerging results suggest that a regression model using low-, mid-, and high-frequency contrast sensitivity and visual lobe size yields a multiple R of roughly 0.8. The predictability is most pronounced for the upper and lower quartiles.

We note that visual lobe size measured by the VLT produces an unusual dispersion of scores. The run of subjects produces dimensions that are approximately normal in distribution; a few Observers have an extremely narrow range of working field of view, producing a somewhat leptokurtotic probability curve centered on the fovea. Sometimes this "tunnel vision" effect is associated with strong corrective lenses, sometimes with no identifiable factor. Such Observers were uniformly poor at target detection.

A very few Observers, however, had virtually flat probabilities of detection within the measured range ($\times 7$ degrees), with detection probabilities in excess of 0.8 across the working field; the VLT did not provide broad enough coverage to define the tails of the distribution. These Observers, not surprisingly, scored exceptionally well on the detection tasks.

Such Observers also tended to have higher contrast sensitivity (and higher Snellen scores), though the relationship between visual lobe size and acuity was less pronounced for normal and low performers. Further investigation of this apparently selective condition must follow.

Another peculiarity of the high performers became apparent in the inspection of eye movement patterns. As expected, Observers with the largest visual lobes required fewer fixations to detect, but it was also apparent that detections were made without fixating on the target itself (by "fixation" we imply centering of the target on the fovea). Figure 3 shows a typical search experience for an Observer with the "super" visual lobe. The mean number of fixations for this observer was four, against a mean for all Observers (out of 21 evaluated at this writing) of 6.6; but the mean of four includes one outlier of 12, in excess of three standard deviations from the Observer's mean; if this maverick trial is discounted, the Observer's mean is 2.8 fixations to detection.

Note also that the target (+) was not specifically fixated; the second saccade passed over it immediately before the trial was halted. Since we may presume saccadic suppression, it seems an inescapable conclusion that the actual detection was made on the third fixation, when the fovea was more than eight degrees from the target.

Figure 3: Example "super lobe" scan pattern. Target was at grid D2/D3 (+)

In essence, these peculiar Observers seem to have a capacity for foveal performance out to the limits of the visual lobe; their probability distributions across the working field of view are remarkably platykurtotic, a condition markedly atypical of the vast majority of Observers tested.

Another possible interpretation -- an attractive one, since the retinal anatomy does not make foveal resolution across 14 degrees a very likely solution -- is that ballistic eye movements in these observers are extremely rapid from programming to execution. If this were the case, then the Observer's eye might be able to dart from the fixation point to the target letter in time to see at least the fading trace of phosphor. In a subsequent experiment we will use a tachistoscopic presentation to test this possibility.

Whatever the origin of these performance differences, the strong predictive model, based on tests that require no more than a wall chart in one case and a microcomputer with millisecond timer in the other provide the possibility for identifying the probable best and worst candidates for tasks requiring high levels of visual target discrimination, and may provide a partial solution for the growing gap between human capacities and weapon performance.

References

Bellamy, L. and Courtney, A (1981). Development of a search task for the measurement of visual acuity. Ergonomics, 24, 497-509.

Bloomfield, J. (1975). Theoretical approaches to visual research. In Fox, J., and Drury, C. (Eds.), Human Reliability and Quality Control (pp. 19-29), London: Taylor and Francis, 1975.

Engel, F. Visual conspicuity as an external determinant of eye movements and selective attention. Unpublished doctoral dissertation, T. H. Eindhoven, 1976.

Gallwey, T. (1982). Selection tasks for visual inspection on a multiple fault type task. Ergonomics, 37(11), 1077-1092.
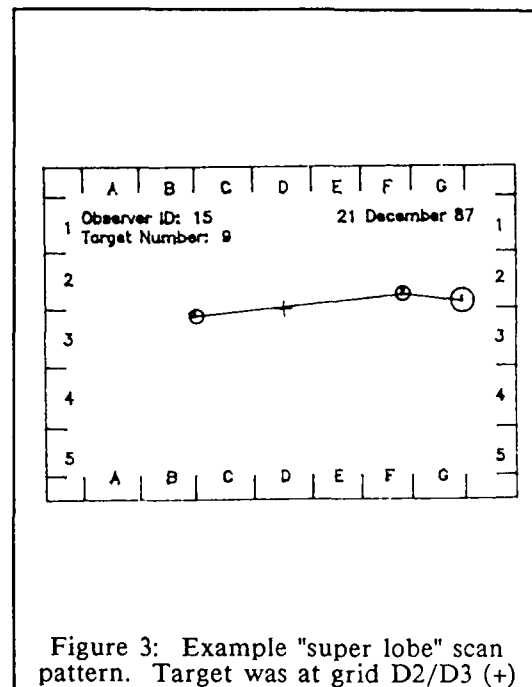
Ginsburg, A. and Evans, D. (1982) Contrast sensitivity predicts pilots' performance in aircraft simulators. American Journal of Optometry and Physiological Optics, 59(1), 105-109.

O'Neill, T. (1986). What the gunner's eye tells the gunner's brain I: Correlates of camouflage detection. United States Military Academy, Human Sciences Report 86-1.

Widdell, H. and Kastner, J. (1981). Eye movement measurements in assessment and training of visual performance. In Drury, C. and Fox, J. (Eds.), Manned Systems Design: Methods, Equipment and Applications (pp. 251-270). New York: Plenum.

USAF Pilot Selection and Classification Systems

Thomas R. Carretta, Ph.D

Air Force Human Resources Laboratory

Abstract

Four hundred seventy eight (478) pilot candidates were given a computerized test battery, the Basic Attributes Tests (BAT) currently being validated for pilot selection and classification. The battery included tests of psychomotor and cognitive/perceptual abilities and personality/attitudinal characteristics. Results indicated that several of the BAT tests were able to improve the prediction of graduation/elimination from flight training and follow-on training recommendation (fighter or non-fighter aircraft) above that provided by currently used paper-and-pencil tests. Implications for pilot selection and classification procedures are discussed.

Until recently, the Air Force has been concerned primarily with the development and validation of methodologies that would screen pilot candidates as early as possible (either by a selection board or during a pre-training flight screening program). Once an applicant has been selected to enter Undergraduate Pilot Training (UPT), the focus shifts from screening to training with the eventual goal of placement in a specialized follow-on training track (either a fighter or non-fighter aircraft).

The current UPT program lasts 49 weeks and involves a T-37 (20 weeks) and a T-38 (29 weeks) phase. The post-UPT training recommendation is made in the 42nd week by an Advanced Training Recommendation Board (ATRB) consisting of T-38 Instructor Pilots. In order for a pilot candidate to receive a fighter assignment he/she must be a fighter volunteer and be recommended by the ATRB. In 1991, the Air Force is planning to adopt a Specialized Undergraduate Pilot Training (SUPT) program. Under SUPT the specialized training assignment will be made prior to entry into flying training. Therefore, it will be necessary to classify pilot candidates into specialized training tracks without the benefit of flying training performance measures.

Preliminary results from a computerized test battery known as the Basic Attributes Tests or BAT, suggest that individual differences in performance on tests of psychomotor skills, cognitive/perceptual abilities and personality/attitudinal characteristics are related to flight training performance (graduation or elimination; fighter or non-fighter assignment, see Carretta, 1987b). In order to be useful as a selection and classification instrument, performance on the BAT battery should add significantly to the validity of the Air Force Officer Qualifying Test (AFOQT), which is being used currently for predicting training success. In particular, it is expected that subjects with better psychomotor skills, quicker reaction times and more efficient memories will be more likely to succeed in training.

Method

Subjects

The subjects in this study were 478 U.S. Air Force officer candidates from the Air Force Reserve Officer Training Corps (AFROTC) and Officer Training School (OTS) who were tested on the Basic Attributes Tests. These subjects already had been chosen for UPT, in part, on the basis of their AFOQT scores.

Instrumentation

AFOQT. The AFOQT, a paper-and-pencil battery, consists of 16 subtests and produces five composite scores: Verbal, Quantitative, Academic Aptitude (Verbal and Quantitative combined), Pilot and Navigator-Technical. Of these, only the Pilot and Navigator-Technical composites are used in the operational selection of pilot candidates (Air Force Regulation 51-4, 26 July 1983, Application Procedure for UPT, UPTH and UNT). The Pilot composite score is based on eight of the 16 AFOQT subtests, including Verbal Analogies, Mechanical Comprehension, Electrical Maze, Scale Reading, Instrument Comprehension, Block Counting, Table Reading and Aviation Information. The Navigator-Technical composite also includes the subtests of Scale and Table Reading and Mechanical Comprehension, but is oriented more toward mathematical and spatial abilities than the Pilot composite. The minimum percentile scores required for entry into UPT were the 25th percentile on the Pilot composite, the 10th percentile on the Navigator-Technical composite and the 50th percentile on the combined total for the Pilot and Navigator-Technical composites.

Of the 478 subjects in this study, 78 completed the AFOQT more than once (2 times, N = 68; 3 times, N = 10). Scores from the first test administration were used for all subjects in the analyses discussed in this paper.

Basic Attributes Tests. This test battery consisted of 12 computer-administered tests that measured psychomotor skills as well as a variety of cognitive and perceptual abilities and personality and attitudinal characteristics. See Table 1 for a brief summary of the BAT battery. For a more detailed description see Carretta (1987a).

UPT Performance Criteria. UPT final training outcome was scored as a dichotomous variable with pass = 1 and fail = 0. Students who passed UPT were evaluated by an Advanced Training Recommendation Board (ATRB) for an advanced training recommendation leading to an assignment either as a Tanker-Transport-Bomber (TTB) pilot or as a Fighter-Attack-Reconnaissance (FAR) pilot (TTB = 0 and FAR = 1).

Apparatus

The Basic Attributes Test (BAT) apparatus was based on a super-microcomputer built into a self-contained unit with a glare shield and side panels designed to ensure consistency of testing sessions. The subject responded to the tests by using in combination or individually a two-axis joystick on the right side of the apparatus, a single-axis joystick on the left side, and a keypad in the center of the test unit. The keypad included the numbers 0 to 9, an "ENABLE" Key in the center, and a bottom row with "YES" and "NO" Keys and two others labeled "S/L" (for same/left responses) and "D/R" (for different/ right responses).

Procedure

Prior to entry into Undergraduate Pilot Training (UPT) each subject was administered both the AFOQT and the BAT. Pilot candidates were commissioned through either AFROTC or OTS. Candidates commissioned through AFROTC took the AFOQT prior to entering college or while an undergraduate. For AFROTC candidates, the BAT was administered during the summer following their junior year in college. For the OTS pilot candidates, the AFOQT was administered after their attainment of a college degree and the BAT was administered at the beginning of their participation in a Flight Screening Program (FSP).

The BAT battery as used in this study consisted of 12 tests lasting about three and one half hours. After the test administrator initiated the system, the test session was self-paced by the subject. The test session included programmed breaks between tests to avoid problems with mental and physical fatigue.

All pilot candidates took part in the same UPT program lasting 49 weeks. The ATRB

decision (TTB/FAR) was made at the 42nd week of UPT, with final training outcome (graduation/elimination) assigned at the end of the program.

Results and Discussion

Three models were evaluated against UPT final outcome and ATRB recommendation. Model I included AFOQT Pilot and Navigator-Technical composite percentile scores along with the number of times the AFOQT was taken by a subject. Scores from individual BAT tests represented Model 2 and the three AFOQT scores combined with the BAT test scores were treated as Model 3.

As shown in Table 2, Model 1 (AFOQT scores) was related statistically to both UPT final training outcome ($R = .169$, $p \leq .01$) and advanced training recommendation ($R = .172$, $p \leq .05$). Subjects with higher AFOQT scores who took the AFOQT only once were

Table 1. Basic Attributes Tests (BAT) Battery Summary

	Test Name	Length (min)	Attributes Measured
1.	Test Battery Introduction	15	Biographical Information
2.	Psychomotor Tests		
	a. Two-hand coordination (rotary pursuit)	10	Tracking and time-sharing ability in pursuit
	b. Complex coordination (stick and rudder)	10	Compensatory tracking involving multiple-axes
3.	Dot Estimation	6	Impulsiveness/decisiveness
4.	Digit Memory	5	Perceptual speed
5.	Encoding Speed	20	Verbal classification
6.	Mental Rotation	25	Spatial transformation and classification
7.	Item Recognition	20	Short-term memory, storage, search, and comparison
8.	Risk Taking	10	Risk taking
9.	Embedded Figues	15	Field dependence/independence
10.	Time Sharing	30	Higher order tracking ability, learining rate and time-sharing ability
11.	Self-Crediting Word Knowledge	10	Self-assessment ability, self-confidence
12.	Activities Interest Inventory	10	Survival attitudes

more likely to complete UPT successfully and to be recommended for follow-on training with fast-jets (FAR).

Of the BAT tests (Model 2), tracking error scores from the two psychomotor skills tests demonstrated the strongest relationship to graduation/elimination from training ($R = .256$, $p \leq .0001$). Psychomotor test performance was related only marginally to advanced training recommendation ($R = .164$, $p \leq .10$). The lower predictive utility of psychomotor performance for advanced training recommendation probably occurred because subjects with poor psychomotor ability tended to be screened out early in UPT (before the ATRB made its recommendations).

Scores from several of the cognitive/perceptual abilities tests also were related statistically to UPT final outcome, but not as closely as the Psychomotor Skills tests. Further, scores on these tests were related to advanced training recommendation.

Subjects who made quick, consistent and accurate judgments were more likely to complete pilot training successfully and to be recommended for advanced training with fast-jets (FAR).

As a group, scores on the personality/attitudinal tests were not related closely to training performance. However, subjects who were more cautious on the test of self-confidence (Self-Crediting Word Knowledge) and chose fewer high risk activities (Activities Interest Inventory) were more likely to complete UPT successfully. None of the personality/attitudinal tests was related to advanced training recommendation.

Model 3 combined 42 scores from the AFOQT and BAT models. Although this model demonstrated the highest predictive validity against both flying training performance measures (UPT, $R = .498$, $p \leq .0001$; ATRB, $R = .435$, $p \leq .05$), it was clear that several of the 42 scores were not contributing to the prediction of training performance.

Table 2. Summary of UPT and ATRB Outcome Regression Analyses

Test	Number of Scores	UPT (pass/fail) M = 0.57 (N = 478)	ATRB (TTB/FAR) M = 0.56 (N = 308)
AFOQT	3	.169**	.172*
Psychomotor Tests	4	.256****	.164
Dot Estimation	5	.062	.139
Digit Memory	3	.108	.108
Encoding Speed	3	.214****	.144
Mental Rotation	3	.156**	.225***
Item Recognition	3	.192***	.211**
Risk Taking	4	.110	.087
Embedded Figures	3	.065	.102
Time Sharing	5	.133	.266***
Self-Crediting Word Knowledge	4	.171**	.105
Activities Interest Inventory	2	.123*	.101
Combined Model (full)	42	.498****	.435*
Combined Model (stepwise)		.436****	.312****

* $p \leq .05$ ** $p \leq .01$ *** $p \leq .001$ **** $p \leq .0001$

In order to identify a less redundant, simpler model, the AFOQT scores were entered first. The remaining 39 variables were allowed to enter the prediction equation in a stepwise manner. Only variables that added to the predictive validity of those already in the model at the .05 level were allowed to enter. This method resulted in a UPT outcome model with 11 variables from eight tests ($R = .436$, $p \leq .0001$). In addition to the AFOQT, variables were entered from both psychomotor tests, three of the cognitive/perceptual abilities tests (Encoding Speed, Item Recognition and Time Sharing) and two of the personality/attitudinal tests (Self-Crediting Word Knowledge and Activities Interest Inventory). The stepwise ATRB model used one score from each of three BAT tests in addition to the AFOQT scores ($R = .312$, $p \leq .0001$). These included Item Recognition, Time Sharing and Mental Rotation. For both outcome measures, the stepwise models did not differ in predictive utility from the 42 variable combined model. See Table 2 for a summary of these regression analyses.

Implications for Selection and Classification

As previously discussed, the Specialized Undergraduate Pilot Training (SUPT) program being considered by the Air Force requires an advanced training assignment to be made prior to the beginning of flight training (UPT). Results from the AFOQT and BAT tests

suggest that uncertainty in making early selection and classification decisions can be reduced by taking into account psychomotor skills, cognitive/perceptual abilities, and personality/ attitudinal characteristics.

These tests could be used in a two step approach. Step 1 (selection) would first screen pilot training applicants for medical problems. Those who were medically qualified would be rank-ordered by their probability of successfully completing UPT. Some cutoff would be established (e.g. 10th or 20th percentile) below which all applicants would be rejected. Step 1 could occur either as part of or after the OTS and ROTC pilot selection board process. If it is to be part of the selection board process, BAT testing could be done at either a centralized location or at several AFOQT testing centers around the country (Military Enlistment Processing Stations [MEPS] and ROTC detachments). If testing is to be done after the board has made a first cut, BAT testing could be done as part of the Flight Screening Program (FSP) or during OTS.

Step 2 (classification) would occur directly prior to the beginning of UPT. The candidates who were selected during Step 1 would be ranked according to their probability of performing successfully in a particular advanced training assignment (fighter or non-fighter). Candidates who are indentified as "fighter qualified" by their test performance and are "fighter volunteers" could be required to meet additional requirements (e.g. pass a G-tolerance test, see Gillingham, 1986). If a fighter candidate fails these additional requirements, he/she could be given a non-fighter assignment.

Summary

The psychomotor tests of the BAT battery are currently under review as adjuncts to the selection process of USAF UPT students and are being used for operational selection by the Air National Guard. In this study, several cognitive/perceptual and personality/attitudinal tests demonstrated potential for improving selection and supporting early prediction of specialized training assignment.

References

Carretta, T.R. (1987a). Basic Attributes Tests (BAT) System: Development of an automated test battery for pilot selection. AFHRL-TR-87-9, Brooks AFB, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.

Carretta, T.R. (1987b). The Basic Attributes Tests: An experimental selection and classification instrument for U.S. Air Force pilot candidates. Proceedings of the Fourth International Symposium on Aviation Psychology, Columbus, OH; The Ohio State University.

Gillingham, K.K. (1986). G-tolerance standards for aircrew training and selection. USAFSAM TR-86-12, Brooks AFB, TX: Aerospace Medical Division, USAF School of Aerospace Medicine.

United States Air Force (1983, July). Application procedure for UPT, UPTH and UNT. (Air Force Regulation 51-4). Washington, D.C., Department of the Air Force.

The   Development of Quick Score Composites for the   Air
Force Officer Qualifying Test (AFOQT) Form P

Todd C. Sperl
Air Force Human Resources Laboratory
Kurt Steuck
Air Force Human Resources Laboratory

Abstract

The   purpose   of this study was to   develop   an
instrument   that   would   enable   recruiters   to
efficiently   identify candidates likely to   succeed
on   the   Air Force Officer Qualifying Test   (AFOQT)
Form P.   The subjects in the study consisted of
roughly   6,000   examinees,   from   the   Initial
Operational   Test   and   Evaluation   (IOT&E)   study
initiated   from   June '87 to Oct '87.   Items   were
selected   from the AFOQT to make   up   subcomposites
referred to as Quick Score Composites (QSCs).   The
QSCs   correspond   to   the   five   composites (Pilot,
Navigator-Technical, Academic Aptitude, Verbal, And
Quantitative) that comprise the AFOQT.   This study
will   result   in   the   development   of
expectancy/conversion   tables   indicating   the
examinee's probability of success on the AFOQT.

The Quick Score Composites (QSCs) are used as a tool to
assist   recruiters   in   identifying   candidates   likely   to
succeed   on   the Air Force Officer Qualifying Test   (AFOQT).
The   QSCs   are   not based on a   separate   screening   device,
rather,   they   consist   of a subset of items   in   the   AFOQT.
QSCs have been in use for Form O since 1982.   New Forms   of
the AFOQT (Forms P1 & P2) were implemented in 1987, but QSCs
have   not yet been constructed.   This paper will discuss the
methods   used   to   build the QSCs for AFOQT Forms P1 &   P2.
This study will be completed by April 1988.

The AFOQT is an aptitude test used by the Air Force   in
the   selection   of individuals into Officer Training   School
(OTS) and Air Force Reserve Officer Training School (AFROTC)
(Wegner   &   Short,   1986).   It   is   also   used   in   the
classification   of   rated   officers   (i.e.,   pilots   and
navigator).   The AFOQT contains 380 items and is composed of
16   subtests   which   are   variously   combined   into   five
composites: Pilot, Navigator-Technical, Academic Aptitude,
Verbal, and Quantitative.

145

The QSCs are needed because of the time delay (often one to two weeks) between the time a candidate takes the AFOQT and his score is reported back to the recruiter. In some cases recruiters need to know how an individual scores shortly after the individual takes the AFOQT. The QSCs are easily scored approximations of the total AFOQT. QSCs exist for Form O, but not for the recently implemented Forms P1 & P2. Forms P became operational in June '87 at non-AFROTC testing sites and in Aug '87 at AFROTC testing sites.

In the past, several prescreening devices have been associated with the AFOQT. Early screening composites were separate tests containing items of the same types as those in the corresponding AFOQT. The Air Force Precommissioning Screening Test (AFPST) was developed to screen applicants for navigator training and to select applicants for the Air Force Academy Preparatory School (Valentine 1961). A preliminary form of AFPST-62 was implemented during Mar '61. A revised AFPST-62, named the Pre-Enrollment Test (PET) was implemented in 1965 (Miller, 1966). The PET was intended as a short screening device for freshman applicants for the AFROTC program. A revision of the PET in was implemented 1967 (Miller, 1968). Items for the PET were selected on item difficulty and internal consistency. Reliability data for the PET-68 subtests and total score were estimated from the PET-66 reliability data. The PET was discontinued in April 1969, leaving recruiters without a prescreening device for officer applicants until one was developed for Form O.

Form O Quick Score measures were developed to correspond with each AFOQT composite by selecting items from each composite (Rogers, 1985). Items to compose the QSCs were selected principally by their high correlations with corresponding AFOQT composite raw scores. Once developed, QSCs were correlated with full-scale AFOQT results across a group of 37,409 AFOQT examinees. A Pearson product-moment correlation was computed between each of the five AFOQT composites and the corresponding QS for the total group (Wegner & Short, 1986).

There are five QSCs for AFOQT Form O, one corresponding to each composite. Each QS is made up of a subset of items contained in the complete composite. The QSCs for the Pilot, Navigator - Technical, Academic Aptitude, Verbal, and Quantitative composites consist of 40, 60, 40, 20, and 20 items respectively.

In summary this study will evaluate several item selection methods and select the technique with the best predictive accuracy. The goal of this study is to provide a product which will allow Recruiting Services to predict, with some accuracy, an examinee's success on the AFOQT Forms P.

Method

The following sections will elaborate on the steps taken in this study: sample and item selection, the construction and evaluation of QSCs, the cross-validation of the selected QSCs, and the development of expectancy tables.

Sample

A developmental sample of roughly 6,000 examinees, from the Initial Operational Test and Evaluation (IOT&E) study initiated from June '87 to Oct '87, will be used. Each version of Form P will consist of approximately 3,000 examinees and will be randomly divided into two groups. This will produce a total of four samples with roughly 1,500 examinees each. One set will be used for item selection and QSC construction while the other set will be used for cross-validation.

Item Selection

One of four methods of item selection will be used to develop a Form P Quick Score. The first method is the selection of items based on the correlations of item performance with subtest scores. Correlations of item performance with composite scores is the basis for the second method of item selection. Methods 3 & 4 will select items which produce the greatest amount of unique variance by subtest and by composite respectively (Gould & Christal, 1976).

QSC Construction

From the four methods of item selection mentioned above a number of alternative composites will be developed and computed in trying to select the QSCs with the best predictive accuracy. During the development of the QSCs two issues need to be discussed. The representation of subtests in the composites and the length of the QSCs vs. the accepted predictive accuracy. First, do we want to keep a similar representation of subtests in each composite, i.e., should the percentage of items selected be proportional to the number of items in the full length composites? Presently

Form O QSCs do not have proportional representation of subtests. Do we want Form P QSCs to mirror the full AFOQT composites? A second issue is the number of items to select. One of the biggest complaints about Form O Quick Score was the amount of time it took to score. However, as the number of items decreases, predictive accuracy will undoubtedly decrease. A number of different QSCs (i.e., 20 to 30 items in length) will be constructed to find the best trade-off between number of items and predictive accuracy.

QSC Evaluation

The alternatives from the QSC construction will be evaluated. A Pearson product-moment correlation between the QSC raw scores and their corresponding full AFOQT composite raw scores will be computed and reviewed.

QSC Cross-Validation

The QSCs with the best predictive accuracy from the evaluation stage will be selected and applied to the remaining sample. Regression analyses and chi-square tests will be conducted on the selected QSCs. The QSCs distributions for various sample subgroups (i.e., gender & ethnicity) should be similar to their respective Full AFOQT Form P distributions.

Based on the evaluation of predictive accuracy in the cross-validation and the assessment of the distributions of subgroups the best QSCs for each of the five Full AFOQT composites will be selected.

QSC Table Construction

AFHRL will then generate expectancy tables which will provide the recruiter the test taker's expected Full AFOQT Form P score. The recruiter will receive the examinees QSC score from the Test Control Officers (TCOs) who score the items and compute the QSC scores.

Results

TCOs will be supplied with 2 items: work sheets and instruction manuals on how to score the QSCs. They will send the computed QSCs to recruiters. The recruiters will be supplied with expectancy/conversion tables and manuals on how to use and interpret the tables.

References

Gould, R. B., & Christal, R. E. (1976). VARSEL: Variable selection for multiple purpose prediction systems in the absence of external criteria (AFHRL-TR-76-6). Lackland AFB, TX: Occupational & Manpower Research Division.

Miller, R. E. (1966). Development of officer selection and classification tests - 1966 (PRL-TR -66-5, AD- 639 237). Lackland AFB, TX: Air Force Human Resources Laboratory, Aerospace Medical Division.

Miller, R. E. (1968). Development of officer selection and classification tests - 1968 (AFHRL-TR-68-104, AD-679 989). Lackland AFB, TX: Air Force Human Resources Laboratory, Personnel Research Division.

Rogers, D. L. (1985). Screening composites for air force officers (AFHRL-TP-85-2). Brooks AFB, TX: Air Force Human Resources Laboratory, Personnel Division.

Valentine, L.D., Jr. (1961). Development of the air force precommission screening test-62 (ASD-TN-61-146, AD-269 527). Lackland AFB, TX: Personnel Laboratory, Aeronautical Systems Division.

Wegner, T. G., & Short, L. O. (1986, November). Assessing the accuracy of the AFOQT quick score procedure. Paper presented at the Military Testing Association, Mystic, CT.

Intelligent Training Worlds:
The FORMS Project

J. Wesley Regian, Ph.D.
Lieutenant Charles G. Capps
Captain Kevin B. Kline
Eliot Shindler
Air Force Human Resources Laboratory

Abstract

This paper describes an ongoing research effort at
the Air Force Human Resources Laboratory (AFHRL)
involving a comparison of a computer-assisted
instruction (CAI) system and an intelligent computer-
assisted instruction (ICAI) system. This comparison is
accomplished by training personnel on a Department of
Defense form, the travel voucher. The FORMS project is
described along with plans for a formal evaluation of
the systems.

Computer-assisted instruction (CAI) is a mature technology
used to teach students in a wide variety of domains. Introducing
Artificial Intelligence (AI) technology to the field of CAI has
prompted research and development efforts in an area known as
intelligent computer-assisted instruction (ICAI). "With the
advent of powerful, inexpensive school computers, ICAI is
emerging as a potential rival to CAI." (Dede & Swigger, 1987)
The key difference between the two is intelligence. There is no
commonly accepted definition of what constitutes an intelligent
instructional system (VanLehn, 1986). We conceive of computer-
based training (CBT) systems as lying along a continuum which
runs from CAI to ICAI. The incorporation of intelligent routines
in CBT systems is expensive and should only occur when there is
sufficient enhancement of instructional efficiency and
effectiveness to offset the additional development cost. This
paper will address this tradeoff in the context of a current
research effort at the Air Force Human Resources Laboratory
(AFHRL). In this paper, we will describe two CBT systems which
differ in the degree to which their behavior is modified by an
inferred "model of the student's current understanding of the
subject matter." (VanLehn, 1986) The CBT system that is less
intelligent by this definition, we will refer to as CAI.
Similarly, the system that is more intelligent, we will refer to
as ICAI. A description of the current research project (FORMS:
Phase I) will be given. This includes a set of criteria used to
make implementation decisions for both CAI and ICAI systems and
the planned evaluation of the systems. Finally, future directions
for this project will be presented (FORMS: Phase II).

The FORMS Project

The FORMS project is an in-house effort of the Intelligent Systems Branch, Training Systems Division, Air Force Human Resources Laboratory. The goal of FORMS is to develop computerized instruction for completing common Air Force forms. These systems are envisioned as tutors rather than as job aids; that is, users should learn the correct procedure for filling out target forms rather than remain dependent on the system for help in filling out forms. We hope to make the processing of Air Force paperwork more efficient by providing cost-effective and instructionally-effective training to new civilian and military personnel in the Air Force. The form targeted for our Phase I effort is DD Form 1351-2, the Department of Defense (DOD) travel voucher submitted by DOD personnel after temporary duty assignments.

In Phase I of the FORMS project, we are conducting a formal comparison of Computer-Assisted Instruction (CAI), Intelligent Computer-Assisted Instruction (ICAI), and traditional on-the-job training (OJT). We are asking two primary questions. First, how does CBT (CAI or ICAI) compare to OJT in this domain? Second, are the additional development costs incurred in developing ICAI (as compared to CAI) warranted for this domain? The answer to these questions will depend on the cost difference among the methods (How much does the training cost?), the relative instructional efficiency of the methods (How long does the training take?), and the relative instructional effectiveness of the methods (How good is performance after training?). Accordingly, we are developing two distinct CBT systems which train personnel to fill out the same form.

Two FORMS Systems

Students begin their interaction with either of the two systems with a survey inquiring about their experience with the travel voucher form. This initial Experience Survey allows us to assess the comparability of the experimental subjects. In addition, this data is used by the ICAI system (but not the CAI system) in generating instruction. When the initial survey is completed, the student is ready to begin receiving instruction. During the instructional phase, both systems keep records of the entire interaction for future analysis.

The CAI system leads the student through the form in a predetermined sequence of instruction. In the case of student errors, the remediation approach is fixed, and remains fixed even if the same error is repeated. In contrast, the ICAI system begins its sessions with an awareness of the student's prior experience in completing the travel voucher. The system uses this information dynamically to plan the sequence of instruction. Furthermore, the level of remediation the student initially receives when an error is made depends on the student's prior experience. When repeated errors occur, increasingly detailed

instruction is provided. The ICAI system is keeping a record of the responses of the student in a way which allows it to respond usefully to the student based on the accumulated student model (see VanLehn, 1986). Both systems are capable of responding to the students questions. For example, the student may need explanations of codes or what incidental expenses are allowable. In the CAI system, students are limited to questions relative to the block they are currently working in. In the ICAI system, students may ask questions about any block on the form, at any time.

FORMS: The "Intelligence" Criteria

In this section we focus on the ways in which our ICAI system differs from our CAI system. For purposes of this project, the differences in our CAI and ICAI versions fall into three categories. These are 1) local vs. global error-checking rules, 2) standardized vs. individualized instruction, and 3) curriculum vs. student directed.

Local vs. Global Error Checking Rules

The CAI system accesses only local error checking rules. For example, in the itinerary block of the form, the traveler is asked to designate his mode of travel for each leg of the journey by recording a symbol (A = auto, B = bus, etc.) in the appropriate box. The CAI system checks to see that the input symbol is on a list of acceptable symbols. The ICAI system, by contrast, accesses both local and global error checking rules. Thus, in addition to verifying that the traveler used a symbol from the appropriate set (a local check), the ICAI system also verifies that the most recent entry is acceptable in light of all previous entries (global checks).

Standardized vs. Individualized Instruction

The CAI system is instructionally more rigid than the ICAI system. The CAI system uses standard text frames as messages to the student. All students using the system see the same instruction prompts and the same remediation prompts. Instruction prompts guide the student through the process of filling out the form. Remediation prompts are provided to the student when an error is detected. When the CAI system detects an error in a block, it generates a message which covers all error types which it is capable of recognizing for that block. The ICAI system utilizes its global knowledge of all information previously input by the student to construct the instruction prompt. For example, in some cases the correct procedure differs for civilian and military personnel. The ICAI system, since it knows that the current student is a civilian, constructs a message with the appropriate procedure and explains that this is the appropriate procedure because the student is a civilian. With regard to remediation prompts, the ICAI system again uses its entire global database, as well as its knowledge of the specific current error,

to construct an individualized prompt. The CAI system, by comparison, would generate much longer messages for cases where the procedure differs between civilian and military personnel. These CAI messages would include both procedures, instructing the student to determine which was appropriate to them.

Curriculum vs. Student Directed

The CAI system is curriculum-directed in the sense that students are forced to proceed in a predetermined linear sequence and can only ask questions about the block they are currently working in. In other words, if the students are at a point in the curriculum when they are learning about the "cost of lodging" block, they can ask questions about that block, but not about other blocks on the form. In the ICAI system, students are guided through a preferred sequence, but they may choose to ask questions about (or fill in) any block at any time. Thus, the ICAI student has access to the entire computer knowledge base at all times.

Evaluation Methodology

We are planning an external evaluation (see Soloway & Littman, 1986) of the two prototype FORMS systems. In this section we outline the planned evaluation effort.

Eighty subjects will be given a verbal description of DD Form 1351-2, focusing on the utility and rationale of the form. Subjects will then be randomly assigned to one of four groups. Experimental group 1 (n=20) will receive CAI training, experimental group 2 (n=20) will receive ICAI training, control group 1 (n=20) will receive guided on-the-job training, and control group 2 (n=20) will receive responsive on-the-job training.

There are two hypothetical TDY (temporary duty) scenarios for the experiment, each consisting of a single-page description of a TDY assignment, a set of orders, facsimile airline tickets, and expense receipts. The two TDY scenarios will be counterbalanced within groups, so that half of the subjects in each group will receive Scenario 1 for training and Scenario 2 for evaluation. The remaining subjects in each group will receive Scenario 2 for training and Scenario 1 for evaluation.

For the two experimental groups, training will be entirely computer-based. An experimental assistant will be available in the room to answer questions and offer assistance in operating the computers, but will answer no questions nor offer assistance regarding DD Form 1351-2. For control group 1 (guided OJT), on-the-job training will consist of a group exercise where an instructor takes the subjects step-by-step through the process of filling out the form. The instructor will answer questions as they arise. For control group 2 (responsive OJT), on-the-job

training will consist of an available instructor to answer questions as the subjects fill out the form.

After training, all subjects will be given a new TDY scenario and a blank DD Form 1351-2. Subjects will be asked to complete the form as quickly and as accurately as possible. Data collected will include: (1) number and type of individual errors; (2) number of errors by error category; and (3) latency to complete the form.

Summary and Future Directions

Phase I of the FORMS project is intended to demonstrate that CBT is a viable technology for the task of teaching government personnel how to complete complex forms. Additionally, we hope that our data will shed light on the degree of sophistication ("intelligence") needed to make the instruction effective. In Phase II, we plan to utilize our Phase I experience to develop a FORMS authoring shell. In the long term, we see the FORMS project as contributing to an on-line help facility for fully computerized government document handling.

The results of the Phase I evaluation effort will also address a more immediate and general issue. We would like to be able to specify, in a principled manner, the level of sophistication appropriate for CBT given a description of the target task (Regian & Shute, in press). Is it feasible to develop intelligent authoring tools that provide guidance to instructional developers relevant to the needed level of sophistication? Phase I of the FORMS project is one of several related projects which explore this possibility.

References

Dede, C., & Swigger, K. (1987) The evolution of instructional design principles for intelligent computer-assisted instruction. Proceedings of the American Educational Research Association.

Regian, J.W., & Shute, V.J. (in press). Artificial intelligence in training: The evolution of intelligent tutoring systems. Proceedings of the Technology in Training and Education conference. Biloxi, MS.

Soloway, E., & Littman, D. (1986). Evaluating ITSs: The cognitive science perspective. Proceedings of the Research Planning Forum for Intelligent Tutoring Systems. San Antonio, TX: Air Force Human Resources Laboratory.

VanLehn, K. (1986). Student modeling in intelligent teaching systems. Proceedings of the Research Planning Forum for Intelligent Tutoring Systems. San Antonio, TX: Air Force Human Resources Laboratory.

Description of Presentation System for an Automated Maintenance Information System

Jeffery Clay, 1st Lt, USAF
Air Force Human Resources Laboratory
Logistics and Human Factors Division

Abstract

The interface between automated information systems and the users of such systems must be defined in a standard way in order to prevent users from having to learn many different systems. The formats and the basic operations used to navigate through or edit data should be consistent throughout all systems. The Air Force Human Resources Laboratory (AFHRL) has done extensive research into the automation of Technical Orders (TOs) and believes this research is valid for other types of data as well. The data formats and the basic operations used for automated TOs will work just as well with diagnostic data or maintenance data collection systems. A description of this interface had been developed by AFHRL and is outlined in this paper. Any document is made up of a combination of elements from one of 4 basic elements or classes described in this paper. Any screen that appears on a computer is one of these types of screens or one of the subclasses under each class. With all types of screens defined, the user has a consistent interface and will always know how to operate the system they are working with.

The Air Force is currently pursuing many automation efforts for the maintenance environment. These include the Automated Technical Order System, the Core Automated Maintenance System, and the Reliability and Maintainability Information System to name just a few. The AFHRL is also working on a project to provide a single interface between all of these automated systems and the maintenance technician.

The purpose of this paper is to describe the interface between the user and an automated maintenance information system. The primary focus will be on the presentation of data to the user and the types of data presented. Any computer screen presentation is made up of certain types of data. These different types of data occur repeatedly with only the actual contents changing. There may be multiple graphics screens in a data base, but the format and operations are similar for all graphics with the main difference being the actual content of the graphic represented. These different types (or classes) of data will be defined, and some basic rules for the display and interaction with each different class will be given.

At the most basic level, there are two elements that make up a display on a computer screen. These are text and graphics. Any information displayed is a combination of these two basic elements. However, to say this defines an interface is ridiculous. Other elements which are made up of these two basic elements must be defined and rules for these larger elements provided.

We have identified four classes of data to be considered in developing and displaying an automated information system. They are: browsers, information, functions, and structure. Specific rules govern the presentation of information in each class. Each class may be broken down into different subclasses which may have different rules from other subclasses in the same class. The classes of information are divided into subclasses as follows: 1) browser - text, graphic, 2) information - all text, all graphic, all table, and mixed, 3) function - dynamic functions, 4) structure - document structure. These classes and subclasses are shown in Figure 1.
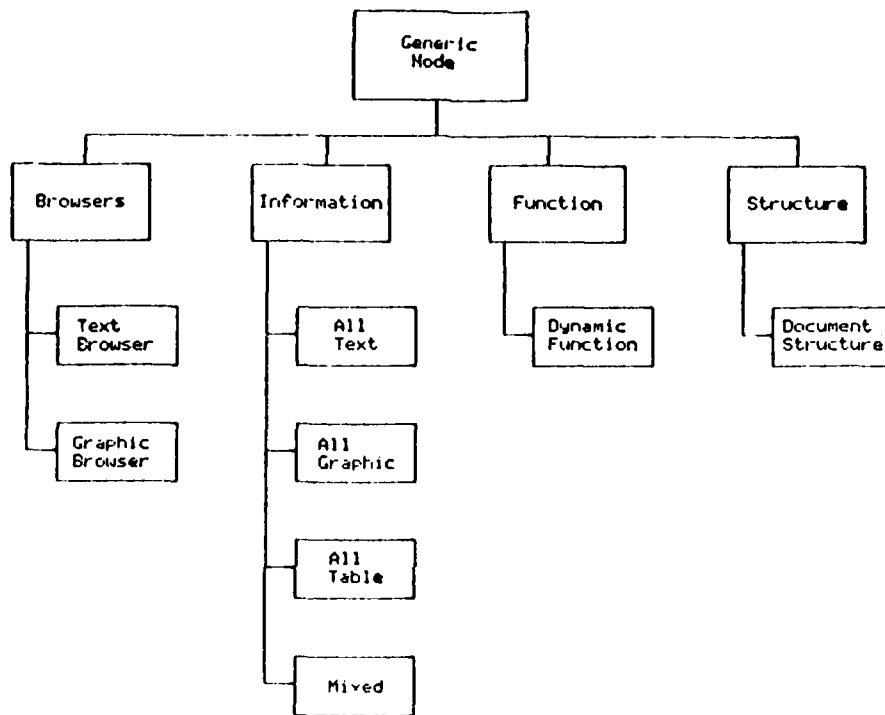
Figure 1. Classes and Subclasses of Information

A browser is defined as a menu or listing of the available information. This is similar to a Table of Contents or a listing of books available in a library. A text browser is a listing of information in text form which may be selected. An example would be a listing of all the procedures in a section or all the chapters in a TO. A graphics browser is an illustration from which certain segments can be chosen. An example would be an illustrated parts breakdown picture or a schematic diagram from which the user can select a part and get more information about that part.

The information class is a presentation that contains technical information for the user. This is by far the most common and has the most subclasses. Information may be text, graphics, or table (combination of text and graphic), or any mixture of the three. There are no pieces to an information frame that can be selected by the user. Data may be entered into an information frame by the user (i.e. entering the results of a test).

A function is a process or algorithm that performs a certain function with data provided by the user. An example of a function would be a diagnostics program that uses test results to determine the next best test for the user to perform.

The structure class defines the allowable structure of a document and contains the title. The document structure for a TO would specify the chapters required to comprise that type of TO. Each chapter would then have a document structure to build each chapter. Structures can exist at almost any level (TO, chapter, procedure, etc.).

Any class may be interactive and require a response from the user. Browsers will always require input from the user to determine the next information to be displayed.

Certain operations will be available to the user regardless of the class of information displayed. These operations include: next, previous, retrace, return, direct access, menu access, references, detail, bookmark, help, and quit. These operations are defined below. Other operations will be available for certain classes of data such as scroll and zoom for graphics.

NEXT - determines the next place to go after the current display is no longer needed.

PREVIOUS - displays the data that precedes the current data displayed (may or may not have been the last data displayed).

RETRACE - displays in reverse order, the data as seen by the user.

RETURN - sets an automatic flag in the data when the user branches to another section and allow the user to return to that specific data without retracing all of their steps (Example: page through several frames of theory and return to the same step in the procedure without retracing each frame of theory that was displayed).

DIRECT ACCESS - allows the user to display any data available without having to go through a series of menus (for use by experienced users).

MENU ACCESS - allows the user to traverse the data by choosing from menus of the data available.

REFERENCES - provides a menu of any reference material that has been linked to current data (schematics, theory of operation, etc.).

DETAIL - allows the user to select a level of detail suitable to their experience level (novice or expert).

BOOKMARK - allows the user to set a bookmark which may be called up at a later time from a menu of bookmarks set by the user.

HELP - provides instructions to the user on how to use the system or information on a specific piece of equipment.

QUIT - allow the user to quit a function or procedure.

The hardware for display of automated data should meet certain
characteristics to be useable. These characteristics are not linked to any
specific hardware, but define the minimum requirements for a usable system.
The response time of an automated system must be no longer than 2 seconds for
most actions. Upon entering an instruction to the computer, unless it
involves uploading or downloading data from another system or other time
consuming activity, the next display should be shown in 2 seconds. If the
action requires a longer response time, the user should get a message telling
how long the action will take. The display resolution must be sufficient to
display complex graphics such as schematics or wiring diagrams in a usable
form. The font used for text should be easily readable from at least arm's
length away for the average user. As a minimum, the keyboard should include a
set of function keys, a set of number keys, a "SELECT" key, an "ENTER" key,
"." and "-" keys, and cursor control keys.
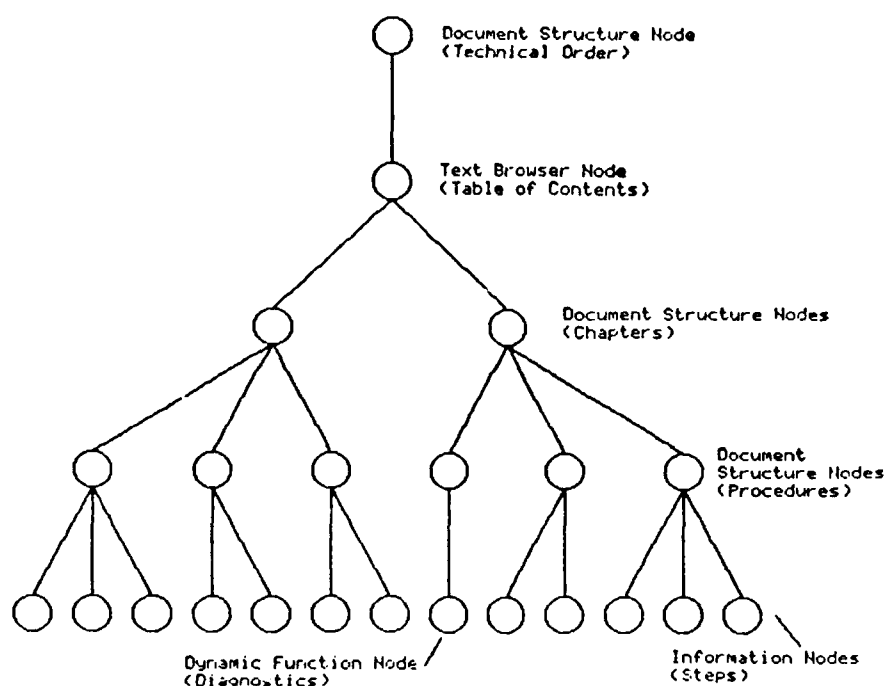
Figure 2. Sample of Automated TO in Hierarchical Format

The subclasses of each data class define the different types of data
allowable to build an automated information system. An automated TO would be
made up of multiple instances of each data class. Figure 2 shows a sample of
what an automated TO may look like in a hierarchical format. The display of
each data class will be in a window in the middle of the screen. The top of
the screen is reserved for location information such as TO title, chapter
title, or procedure title. The appropriate Maintenance Integrated Data Access
System (MIDAS) number would also be displayed here if applicable. The bottom
of the screen is reserved for the available options. These options are dis-
played in a function-key manner corresponding to function keys on the computer.
A sample screen display is shown in Figure 3.

```
┌──────────────────────────────────────────────────────────┐
│ 12P4-2APX64-2          REPLY ENCODING CHECK       2.0.1.1C│
│ a.  Set TRANSPONDER SET CONTROL MODE 1 switches          │
│     to 00 and CHAL/TAG MODE 1 switch to ON.              │
│                                                          │
│ b.  Rotate AN/APM-239A TRANSPONDER SET CONTROL MODE      │
│     1 switches (1) and verify that code pulses           │
│     appear in proper sequence.                           │
│                                                          │
│     For detailed instructions,                           │
│     refer to Theory of Operation                         │
│                                                          │
│                                                          │
│   F1 C1 A1 C2 A2 C4 A4 X  B1 D1 B2 D2 B4 D4 F2           │
│                                          APM-239A        │
│   EXAMPLE:  MODE 1 CODE 73                               │
│                                                          │
│     Do pulses appear in proper sequence?  Yes or No      │
│ [1] YES  [2] NO  [3] LESS  [4] DIRECT [5] T.O.C. [6] THEORY [7] PULSES [8] ETC. │
└──────────────────────────────────────────────────────────┘
```

Figure 3. Sample Screen

Rules are needed to describe how each data class is displayed in the data window. Text browser data would be displayed in a menu format with each entry numbered for selection. Selection is made by entering the number of the desired choice. A graphics browser would be in the form of a graphic with the cursor movement controlling selection. When the cursor is placed over a selectable item on the screen, it is highlighted. Selection is made by pressing the ENTER key. Text information will simply be printed in the data window beginning at the upper left corner. Graphic information will be scaled to fit the data window and scroll and zoom capability will be available if the graphic is larger than the window size. Table data will be displayed in table format with scroll capability if necessary. The table column and row headings will remain visible on the screen even when scrolling to see more of the table. Any mixed data will be displayed using rules to determine the optimum screen layout.

AFHRL has done extensive work on the automation of TOs. This experience includes field tests with prototype automated TOs as well as in-house work developing an Authoring and Presentation System. An object-oriented approach has been chosen to provide flexibility and growth potential. AFHRL is continuing to conduct field tests with the users of an automated information system. The man-machine interface will continue to develop and grow as a result of this on-going research.

Developing a Training Decision Table —
Deciding What Tasks to Train

Michael Fodale
Charles T. Jervey
USAF Occupational Measurement Center
Randolph AFB TX 78150-5000

Abstract

The Air Force Occupational Measurement Center's (USAFOMC)
Occupational Analysis Division (OMY) collects information about
career fields through carefully designed occupational surveys. The
Air Force uses Occupational Survey Report (OSR) results for defining
and maintaining occupational structures within the Air Force
classification system, for adjusting and establishing training pro-
grams, and for sustaining and modifying other Air Force personnel
programs. This paper describes an automated version of the decision
logic contained in the Course Training Decision Table, and Automated
Training Indicators (ATI) derived from computer applications which
enable enhanced use of occupational analysis information by pro-
viding data in a more timely manner and viewed in different formats.

The Air Force Occupational Measurement Center's (USAFOMC) Occupational
Analysis Division (OMY) collects information about enlisted, officer, and on
occasion, Department of the Air Force civilian specialties through carefully
designed occupational surveys. The Air Force uses Occupational Survey Report
(OSR) results for defining and maintaining occupational structures within the
Air Force classification system, for adjusting and establishing training pro-
grams, and for sustaining and modifying other Air Force personnel programs.

Three basic types of information resulting from occupational surveys and
used for making training decisions are: (1) percent members performing—how
many people do a certain task, (2) training emphasis (TE)—which tasks should
be emphasized in structured training for entry-level personnel in the spe-
cialty, and (3) task difficulty (TD)—how hard is each task to learn in com-
parison to other tasks.

Training Emphasis (TE)

TE relates to the rating scale which measures perceptions as to which
tasks should be emphasized in structured training for entry-level personnel.
Structured training includes basic resident technical training, formal on-
the-job training (OJT), and career development courses. TE ratings inherently
include consideration of the consequences of inadequate performance and task
delay factors. TE data are collected from experienced field personnel and are

rated on a scale of zero (no training required) to nine (extremely high training required). TE ratings are used to guide training decisions.

Task Difficulty (TD)

TD relates to a rating scale which measures the relative difficulty of tasks performed by a specialty, where difficulty is defined as the length of time required by an average incumbent to learn to do the task. These ratings are also collected from experienced field personnel during the survey process. Because all tasks have some degree of difficulty, TD is rated on a scale from one to nine; one being extremely low in difficulty, nine being extremely high in difficulty. This rating scale allows for rank-ordering of tasks by relative difficulty through a process that transforms to standard scores to establish an average or mean. In this regard, TD ratings are adjusted so that tasks of average difficulty have a mean rating of 5.00 and a standard deviation (SD) of 1.00. TD ratings are used to identify the tasks which take longer to train than others and to guide training decisions.

Training Decisions

Technical training school personnel use task factor data to make several types of training decisions. Some of the most common decisions to be made are whether or not to change the length of the basic course, to add or delete instruction on various equipment, or how to rewrite training documents. One tool training personnel have for making these decisions is a Utilization and Training Workshop (U&TW). The basic concept of the U&TW is to review how personnel are being used, determine how they should be used, and decide on the training program, both formal and OJT, to minimize the difference. During the course of a U&TW, results of the OSR for that particular specialty are briefed, and AFMPC briefs on the career ladder structure and reviews current personnel utilization, authorizations, manning, etc. MAJCOMs discuss their training programs, to include OJT, and the training manager (TM) reviews career ladder training programs to include the basic, supplemental, special, and career development courses. The TM briefs on the Specialty and Course Training Standards (STS/CTS) and their impact on training. Finally, AFRs 36-1 and 39-1 are reviewed, as well as the training standards supporting them (AFR 8-13, ATC Sup 1, 1987).

ATCR 52-22, Occupational Analysis Program, provides general guidance on use of survey data in reviewing STSs and entry-level Plans of Instruction (POI) for courses created to support specialty training needs. A Course Training Decision Table assists training personnel in determining what type of training is needed and to what depth. Type of training recommended may be centralized training such as an Airman Basic Residence (ABR) course, Field Training Detachment (FTD) course, or decentralized training, such as OJT. Depth of training recommended may be either principles and performance, principles only, mentioned as a brief teaching step with most training accomplished through OJT, or through OJT only.

Automated Training Indicators (ATI)

In the Fall of 1986, an automated version of the decision logic contained in the Course Training Decision Table was developed. These Automated Training

Indicators (ATI) are now available to training personnel in several useful formats. ATI were developed through the use of a series of Comprehensive Occupational Data Analysis Programs (CODAP) which processes OSR data and develops numeric indicators corresponding to the decisions of the Course Training Decision Table. Otherwise, ATI are derived from computerized application of criterion group percent members performing, TE data, and TD data, using the logic given in the table. They're normally presented in descending order from highest ATI (18) to lowest ATI (1).

The Course Training Decision Table is made up of five columns. The first column references the criterion group, normally first-enlistment personnel (1-48 months TAFMS), and addresses the percent members performing a given task. The second column addresses TE ratings, either high, above average, or average or below average. In the third column, TD ratings are addressed. The fourth column addresses whether or not a given task is critical or involves a safety issue. The fifth column gives the resultant training decision dependent upon which blocks are addressed across columns one through four.

There are 18 decisions based on four categories: (1) train at task knowledge and performance levels or at task knowledge only; (2) train at OJT unless centralized training is justified; (3) train at OJT unless justified for criticality or safety reasons; and (4) train at OJT only. In the ATI table, ATI numbers appear in the fifth column, with the decision wording placed off to the side. This version of the ATI table is seen in USAFOMC training extracts for the early (FIELDATA) and newer (ASCII) versions of CODAP task factor printouts (FACPRTS). A short or condensed version of all ATI decisions is also available for use as a handout at U&TWs along with FACPRTS.

Figure 1 displays a sample page from the AFSC 545X1, Liquid Fuel Systems Maintenance, career ladder STS with ATI. Tasks matched to paragraph 9C, Use Specialized Tools, Protective Clothing, and Equipment to Clean Tanks, are presented with percent members performing for first-enlistment groups, TE and TD data, and ATI. For example, task F126, Ground portable equipment, has 76% members performing, has a TE of 5.21, a TD of 3.74, and an ATI of 18.0. According to the ATI decision table, an ATI of 18.0 requires centralized training of both task knowledge and performance. At the other extreme, task B25, Direct tank cleaning operations, has 10% members performing, a TE of 2.18, a TD of 7.98, but an ATI of only 2.0. According to the ATI decision table, a 2.0 requires no centralized training and can be accomplished through OJT.

Figure 1. Sample data page showing ATI

545X1 STS/PERCENT PERFORMING/TE/TD = ATI

D	TSK	TITLES	1ST ENL	TNG EMP	TSK DIF	ATI
138		9C. USE SPECIALIZED TOOLS, PROTECTIVE B 3C 4D CLOTHING, AND EQUIPMENT TO CLEAN TANKS				
F	126	Ground portable equipment	75.8	5.21	3.74	18.0
G	235	Squeegee tank floors	57.3	4.06	3.69	17.0
G	201	Install or remove blowers or educators	38.2	4.15	4.67	15.0
G	218	Operate manual fresh air blowers	43.3	5.64	4.42	12.0
G	220	Perform tank floor leak checks with vacuum box	4.5	3.85	6.80	7.0
B	25	Direct tank cleaning operations	9.6	2.18	7.98	2.0

162

ATI has been field tested at U&TWs by USAFOMC analysts and training personnel with favorable results. ATI enables enhanced use of occupational analysis information by providing data in a more timely manner and to be viewed in different formats. Meaningful decisions which occur at U&TWs as a result of ATI use are achieved sooner and made with greater confidence. ATI allows training personnel to make data-based decisions, on what type of training will best serve the needs of the Air Force in producing a competent specialist (using current task data on who performs what in a given specialty).

References

Fodale, M. (1987). Automated Training Indicators. Proceedings of the 1988 Technology in Training and Education (TITE) Conference.

Fodale, M. & Aslett, L.S., Ph.D. (1987). Automated Training Indicators. USAFOMC Technical Note Series 87-01. Randolph AFB TX.: Occupational Analysis Division, USAF Occupational Measurement Center, (ATC).

Air Force Regulation (AFR) 8-13 (1986), ATC Supplement 1 (1987). Air Force Specialty Training Standards and Air Force Job Qualification Standards. Department of the Air Force, Headquarters Air Training Command, Randolph AFB TX.

Cognitive Processes in Question Asking and Their Use for Developing
Computer-Based Interactive Tutorial Dialogues

Merryanna L. Swartz, M.S.
U.S. Army Research Institute
for Behavioral and Social Sciences
Alexandria, Virginia

Abstract

Understanding how soldier trainees ask questions in computer-based learning will be useful for the development of intelligent tutorial dialogue interfaces. This research describes the cognitive role of question asking during skill acquisition in a computer-based learning by doing paradigm. Results indicate that questions are used by students to guide organization of the knowledge, discriminate between states, and obtain instructional information to facilitate repair of a given solution attempt (Brown & Van Lehn, 1980) when problem solving strategies are unsuccessful.

Soldiers will be interacting more directly with high technology computer-based training systems in the Army of the future. Soldier trainees interacting with these machines will profit from the ability to ask questions as they learn. A fundamental problem we need to address is how to create computer-based tutorial dialogues that will allow easy, natural language question asking to help facilitate soldier training on these systems. A first step toward accomplishing this goal is to understand the cognitive role of pedagogical question asking in computer-based learning.

Questions asked during acquisition of a complex skill reflect what information military students require at different stages in learning. Evaluation of the kinds of questions asked in a particular MOS should provide information that is descriptive of each student's current knowledge state. Question analysis can be used for defining the goals, operators, and methods a student is using or needs at a given time during the instructional sequence. This information can be useful for student modeling in intelligent tutoring systems. Understanding the purpose of question asking is an important issue for creating intelligent tutorial dialogues wherein both the tutor and student can ask questions to clarify acquisition of new concepts, guide student performance, and thereby improve both learning and instruction. I argue that student-generated questions function within pedagogy as a metacognitive process to aid the student in understanding new knowledge (Brown & DeLoache, 1978; Bransford, 1979). The ability of a student to articulate questions during learning is based on the notion that knowledge appropriate structures and a certain level of completeness of these structures are available in the student's internal representation (Miyake & Norman, 1979). Furthermore, there is evidence (Chi & Glaser, 1980; Swartz, in press) that novices do in fact ask many questions when learning a new skill, even if these questions are ill-formed. The cognitive process underlying the proposed model of pedagogical question asking are based on the theory of spreading activation in a semantic network representation for the knowledge. I will argue that student pedagogical questions reflect in part self-directed probes in memory (Brown &

DeLoache, 1978) which function as a control mechanism for focusing activation at a particular knowledge node in order to maintain it in working memory during processing (Swartz, in press). As a second argument, I claim that pedagogical questions function as organizers for the knowledge (Miyake & Norman, 1979) through discrimination and generalization processes as students evaluate what they know with the feedback received from tutorial responses and task goal end states.

The studies presented explore the function of pedagogical question asking at the initial stages of skill acquisition in a learning by doing paradigm. I will use a question interpretation model based on a conceptually dependent representation for the surface structure of a given question to analyze the verbal protocols collected. This method allows a mapping of the question unto a conceptual representation for the domain knowledge. Using this procedure, I will present data from both studies to support the claim that questions are posed in order to organize incomplete knowledge structures that are taking form during learning. I will show that the questions generated at impasse (Brown & Van Lehn, 1980) are used to help guide subsequent problem solving so that future similar experiences will contribute positively to the formulation of the knowledge structure and eventual skilled performance. In the second experiment, I will provide evidence for the changes in type of question asked as a function of practice and feedback to the questions.

Experiment One

Subjects

Six undergraduate college students at the Catholic University of America participated in this study. This experiment was part of their psychology course requirement.

Procedure

The experiment was a within-subjects single group design measuring number and type of questions asked per spreadsheet. The domain under investigation was LOTUS 123, a computer command language. Four simple spreadsheets were prepared. Each one was similar in subgoal structure and length. The task involved reconstructing these spreadsheet problems on a computer. Order of presentation of the spreadsheets was randomized. Each subject was given 15 minutes of tutorial instruction in LOTUS 123. Five minutes of additional practice with sample exercises was conducted before the study began. The trial was conducted on an IBM-XT equipped with a hard disk drive. All sessions were video-taped using VHS video recorder equipment. A keystroke and timing program were used to capture keyboard behavior and time on task. Subjects were instructed to think aloud as they solved the spreadsheet problems and to ask questions aloud when they weren't sure of what to do or how to proceed in the task. Questions were not answered until after the session, however, subjects were encouraged to ask questions anyway and then try to reconstruct the spreadsheets as best as they could. After the session, the experimenter provided them with answers to their questions.

Experiment Two

Subjects

Thirty-two undergraduate college students at the Catholic University of America participated in this study. This experiment was part of their psychology course requirement.

Procedure

The experiment was a three-factor mixed design 2 (practice) X2 (feedback to questions) X3 (spreadsheet-within subjects). The overall procedure in this experiment was similar to experiment one except for the conditions listed below. Each subject was given a 30 minute tutorial. The practice groups received a one hour guided practice session prior to the experimental trial. The feedback group received answers to questions in the experiment. The experiment consisted of three spreadsheet problems.

Results and Discussion

Each question, obtained from verbal protocols collected during the experiments, was analyzed by identifying the interrogative pronoun and placing it into a conceptual category to reflect the predicate relation of the question: causal, procedural, feature specification and so on, to the knowledge asked about. The propositional unit is the remainder of the sentence after the question particle is removed. Next, this propositional unit was divided into two components, the first of which was categorized into a context classification to reflect the instructional problem space a student was in when the question was asked. The second component of the proposition is the unknown element of the question (Graesser & Black, 1985). The knowledge element is information the questioner needs to have to complete understanding. The general frame proposed for analyzing pedagogical questions has the following form: (<Interrogative Pronoun><Problem Space Context, Knowledge Node>)
How do I enter it? => (<Procedural><Action Space, Data Entry Node>)

Subjects in experiment one provided a total of 130 questions. Three main types of questions accounted for more than 70% of the questions asked during the trial, procedural questions (29%), goal questions (22%), causal questions (30%). These results indicate that naive learners need both basic procedural information to perform in the task and task goal information from the onset of learning. The high proportion of causal questions reflect the student's need to understand what is going on and to recognize their own errors. Three kinds of knowledge accounted for 64% of the content requested in the questions, data entry (35%), editing (15%), and calculations (14%). A chi-square test on the data matrix for the most frequently occurring question types (procedural, causal, goal, verification) and knowledge accessed (data entry, editing, calculations) by the questions revealed a significant relationship between procedural questions and the three knowledge nodes, X^2 (2, N=6) = 7.95, p <.05. The relationship between the data entry knowledge node and the three question types approached significance, X^2 (2, N=6) = 5.38, p <.05.

In the second study, a total number of 260 questions were asked. The mean number of questions asked per group was 25 for group 1 (no feedback/no practice), 14 for group 2 (no feedback/practice), 31.3 for group 3 (feedback/no

practice), and 16.9 for group 4 (feedback/practice).A doubly multivariate analysis of variance (MANOVA) repeated measures test measuring number of questions asked and time on task as dependent variables was carried out. A main effect for the number of questions asked in the practice condition approached significance, $F(1, 0, 12 1/2) = 2.98$, $p= .06$, but there was no main effect for the feedback condition ($p <.05$). For time on task, there was a main effect for the practice condition, $F(1, 0, 12 1/2) = 6.15$, $p=.006$, but no main effect for the feedback condition ($p <.05$). The Feedback X Practice and Feedback X Practice X Trial interactions were not significant for either outcome variable. The no practice groups asked more procedural, causal, and verification questions than the practice groups which supports two hypotheses examined in this study. 1) The type of questions asked changes as a function of practice and consequently more developed knowledge representations. To articulate a question, the student must be able to express what it is he or she needs to know. One basic assumption for this ability is that the student has already acquired a certain level of knowledge from which to formulate a certain type of question (Miyake & Norman, 1979). 2) Students use questions to organize the knowledge acquired. Causal questions when state evaluations are being made provides evidence that students may be discriminating between expected and actual system states using question feedback to identify correct states, recognize student errors, and organize and correctly process the knowledge. Also, verification questions can illustrate how novices unsure of procedures may seek confirmation of an action to help them learn correct methods for the skill in a learning by doing environment. Miyake and Norman (1979) suggested that students who are capable of generating and testing hypotheses are in the process of actively constructing some type of organizational structure for the knowledge. The feedback/no practice group generated the most causal and verification questions in state space contexts, evidence for discrimination learning as students attempt to organize and evaluate what was being learned. These kinds of questions were most often asked when students were unable to problem solve successfully. This finding suggests that in the absence of practice, feedback to pedagogical questions facilitates learning. Table 1 lists the distribution of the most frequently asked questions in the second experiment by category (procedural, causal, verification) and problem space context.

The results of these studies indicate that question asking by students is used to monitor the learning process and guide the organization of domain knowledge. Problem solvers attempting to formulate a representation for the task can combine question asking with evaluation of individual performance behavior from previous similar situations to organize and understand task knowledge in a learning by doing paradigm. Student pedagogical questions function in general to identify problems the student has with understanding a particular concept or procedure in the domain. When a student's problem solving does not enable him or her to "repair" (fine a correct solution) at "impasse" (when problem solving does not solve the problem) (Brown & Van Lehn, 1980), an answered question can provide the necessary information for repair strategies. Tutoring systems need to know how much and what kind of information the student knows before an appropriate response can be given (Burton & Brown, 1979). If the tutoring system can interpret the knowledge state from the types of questions a student asks, than an appropriate answer is possible. The descriptive data presented here describe important information for addressing this problem.

Table 1. Mean Number of Questions Asked by Category Problem Space Context and Experimental Groups

Question Category				Problem Space Context		
	Pro	Causal	Verify	Goal	Action	State
No Feedback/ No Practice	20	16	22	15	30	27
No Feedback/ Practice	13	3	15	10	24	8
Feedback/ No Practice	19	13	41	21	45	28
Feedback/ Practice	10	8	27	12	19	21

References

Brown, A. & DeLoache, J. S. (1978). Skills, plans, and self-regulation. In R.S. Siegler, (Ed.). Children's thinking: What develops? Hillsdale, NJ: Lawrence Erlbaum Assoc.

Brown, J.S. & VanLehn, K. (1980). Repair theory: A generative theory of bugs in procedural skills. Cognitive Science, 2, 155-192.

Bransford, J.D. (1979). Human cognition, learning, understanding, and reminding. Belmont, MA: Wadsworth Publishing Co.

Burton, R.R. & Brown, J.S. (1979). Toward a natural-language capability for computer-assisted instruction. In H.O. O'Neill (Ed.). Procedures for instructional systems development, NY: Academic Press.

Chi, M.T.H. & Glaser, R. (1980). The measurement of expertise: Analysis of the development of knowledge and skill as a basis for assessing achievement. In E.L. Baker & E.S. Quellmalz (Eds). Educational testing evaluation. Beverly Hills, CA: Saye Publications.

Graesser, A. & Black, J.B. (1985). The psychology of questions. Hillsdale, NJ: Lawrence Erlbaum Assoc.

Miyake, N. & Norman, D.A. (1979). To ask a question, one must know enough to know what is not known. Journal of Verbal Learning and Verbal Behavior, 18, 357-364.

Swartz, M.L. (in press). A cognitive model of pedagogical question asking. Technical Report, U.S. Army Research Institute, Alexandria, VA.

Organizational Designs for Managing Expert Systems

John R. Landry
Iowa State University
Major Rita A. Moore
USAF Academy

Abstract

Expert systems have generated interest from practitioners and academics who recognize the potential for the development and use of the systems which allow expertise to be applied for solving ill-structured problems. However, little attention has been directed toward the design of organizations to promote effective utilization of this technology. An organizational design model, derived from a transactions cost framework, is proposed to improve organizational effectiveness. The approach considers system characteristics including the certainty of output choices and type of knowledge contained in the system.

The rapidly changing information systems environment is forcing the need to better understand the management of information systems (Zmud, 1984). Traditional management frameworks for information systems technology usually have been founded on technical grounds rather than organizational needs (King, 1983); however, exclusive dependence on technical criteria may overlook fundamental organizational requirements. Weick (1969) underscores this need to consider the organizational psychology of the human interacting with objects as well as humans interacting with each other. "Organization is a mediated causal relationship between two items [humans and/or objects], in which the relationship between the two items is influenced by the third."

Expert systems pose organizational design issues beyond those of typical information systems because they are not only composed of facts and algorithms but include knowledge and heuristics derived from human experts. System outputs typically consist of multiple answers, a list of possible choices which reflect the uncertainty or complexity of the problem space, along with the confidence level of each choice. Since the solution is guided by heuristics and explicit representations of an expert's beliefs and biases in a given domain, these systems do not provide the user with the same predictability of output found in conventional information systems and may require a different management approach.

We will describe an organizational design to improve the organization's performance when expert systems are used. The model provides for various forms of managing the relationship between a human and the expert system by considering: (1) the uncertainty of task performance, and, (2) the type of knowledge contained in the expert system. The remainder of this paper will discuss current management approaches and the transaction cost based model.

Prescriptions for the Management and Control of Information Systems

Prescriptions offered for information systems management are generally founded upon one of two philosophies, ethical or technical. Ethical arguments claim that knowledge is gained by sharing the information available to promote learning and providing the systems to share information. This school of thought believes that information is, or should be, a free good and there is no reason within the organizational to limit access except for security reasons. Technically based systems prescriptions adopt a "rational" model and structure management in a logical fashion so that costs and benefits can be allocated. Furthermore,

rationalists take one of two general positions regarding who should bear the responsibility for information management and control - the creator (source) or the organization's management.

Closely related to the management of knowledge bases is the data administration function. Research in this area which has been limited (Kahn, 1983) and generally posits that control is provided by the organization's existing management through allocation of its ownership rights. Ballou and Tayi (1985) state "At least in theory the organization is the owner." Brathwaite (1983) also asserts ownership belongs to top management to delegate as it sees fit. His prescription to top management is to develop classes of rights to data, and to owners is "not to abuse those rights and block or frustrate other departments' efforts to access the data." This research in data administration provides only limited guidance for the design of management structures for expert systems.

Transactions Cost Paradigm

An alternative to the previously cited management approaches is offered by Williamson's (1981) transactions cost paradigm. It has been described as central to an integrated theory of information systems (Bakos and Treacy, 1986) and as a promising organizational design approach (Jones, 1983). This paradigm views the organization as a set of exchanges or transactions between members to achieve some goal (Jones, 1983; Ouchi, 1979; Williamson, 1981) and structures these transactions contingent upon the nature of the transaction. Contingency approachs to have been adovated by King (1983) who believes that organizational structures must be related to decision making activities. Williamson's approach is significant because it provides specific frameorks for the management of exchanges (tasks) between organizational units (groups or individuals) at the site, physical and human asset level. These exchanges can be physical goods or information. He posits organizational design, or governance structure, follows from the need for organizational units to make exchanges in a manner that allows them to overcome constraints imposed by bounded rationality, uncertainty, opportunism, limited alternatives, assymetric information, and environmental conditions.

Governance structures are "social structures and processes" whose objective is "maintaining the perception of equity among the participants to a transaction." (Ouchi, 1980). Equity is important for the maintenance of continuing exchange relationships (e.g., employment, job performance, etc.). Management and control becomes an issue of finding an appropriate governance structure to insure that the goal directed exchange between the user and the system is efficient and effective (Jones, 1983). Central to designing effective governance structures is evaluation of the exchange on each exchange dimension - transaction frequency, performance uncertainty, and asset specificity (Williamson, 1981).

Transaction frequency is assessed relative to number of times a transaction is conducted over long periods of time. It is the least important of the dimensions because nonrecurring exchanges are typically handled on an ad hoc basis (e.g., task forces). Only in recurring transactions, such as interactions with an installed system, do the participants need to work out a process that will assure equity over long periods.

A second dimension, performance uncertainty of an asset, is related to the user's perception of usefulness of the asset (i.e., expert system) in a specific application. If the performance is relatively certain, then an price may be established which reflects the value of the asset and the use of that asset can be transferred to the user for a price through a transaction protected by established procedures. On the other hand, if the performance can not be determined a priori then some form of procedure must be devised to work out problems arising from unexpected situations encountered during exchange or use.

170

For example, if the performance of an expert system is known with a high degree of certainty then the expert system may be acquired from a source for an equitable price. Any problems arising during the exchange or use may be resolved through established policy. If a greater degree of performance uncertainty is expected then the user wants greater assurance the expert system will be of value and, consequently, desires a greater control over the exchange between himself and the system to insure performance. Operationally, an expert system that provides a small number of options, with high confidence factor for one option, would appear to the user as reducing uncertainty. Of course, the user's trust in any system would only come with use over time, under a variety of conditions.

The third dimension, asset specificity, refers to the uniqueness of the asset to the organization's intended application relative to all other uses. Indiosyncratic assets may contribute to development of sustainable, superior performance and strategic advantage. However, they consume resources and are not easily converted to other resources. Although expert systems are not explicitly discussed within the transactions cost paradigm, this dimension is discussed by drawing an analogy between human asset specificity and expert system knowledge bases. Idiosyncratic human assets arise from learning by doing and having specific organizational knowledge (Williamson, 1981). While learning by doing is a necessary condition, it is not sufficient condition to assure asset specificity - learning must be coupled with specific organization knowledge. In the case of operational expert systems, the knowledge bases developed, or learned, from a generally well understood processes are not sufficient to give rise to asset specificity (e.g., commercial vehicle repair knowledge base). Only if knowledge base contains organizationally specific knowledge (e.g., battlefield tactics) then the expert system is idiosyncratic.

Organizational Design Based on Transaction Cost

Existence of an idiosyncratic asset or performance uncertainty or both indicates an opportunity for exploitation of one or both parties by the other during a exchange. For example, the system could be exploited by the user failing to provide complete information or using the system in inappropriate situations. Similarly, the system could exploit the user through the biases of the expert built into the knowledge base. Another organizational consideration is that once a system is chosen for use, access to alternative systems is limited and the potential effects of exploitation are greater. Under conditions conducive to exploitation then the equity of the exchange must be protected through a more flexible mechanism than a simple exchange of use (knowledge) for a price (e.g., dollars, labor, etc.).

This additional transaction assurance could be arrived at through a policy which specified all possible states of the world with remedies if those states occurred. However, since exploitation, bounded rationality, limited alternatives, and uncertainty exist and limit the ability to develop a comprehensive policy, some form of organizing other than policies is warranted. The alternative is greater reliance on effective management. Four organizational forms derived from the transactions cost model are described below and each form is described relative to appropriateness for given levels of performance uncertainty and knowledge base specificity.

For low system performance uncertainty and low knowledge base specificity, management is via internal market-like mechanisms (e.g., GSA) that allow anyone access to the expert systems for a transfer price between the user and supplier. Expert systems procured for use in this manner allow the user to acquire the system if its benefits are understood and resolve any performance disputes using established policies and procedures. The management of the system is identical to managing any other asset which has well known performance characteristics (e.g., procuring and using a typewriter). Should the item

171

fail to perform for the users the internal market mechanism signals other potential users to change their expections or change the price of the item.

When uncertainty increases and the knowledge base specificity remains low then the expert system is joined with a human to form a team and the teams are assigned a task. Only the total output of the team can be evaluated since it is impossible to decompose the contributions of the each party. Since the specificity is low the items in the team are mobile, that is there is a market for the services beyond the immediate organization. Since this market exists the organization is more able to procure replacements from other sources and is less concerned about retention.

As knowledge base specificity increases and performance uncertainty remains at low levels, then organizations may decompose the work and monitor progress through bureaucracies. This approach is similar to the "rational" control approach which provides for cost benefit analysis of each component. Costs and benefits of the system, operators, maintenance, etc. can be analyzed and appropriate management actions taken on the basis of that analysis.

At high levels of uncertainty coupled with a highly idiosyncratic knowledge base a relational team, or clan, type of management is appropriate. In this form of control the shared values, peer pressure, and diverse viewpoints, including the output developed by the expert system, seek an interpretation of the circumstance. This approach differs from the team in at least two respects. First, the clan members must develop a strong trust relationships between themselves and others. Second, loss of member's idiosyncratic knowledge, including the expert system's knowledge, may result in loss of the organization's strategic position.

The organizational designs to achieve high levels of organizational performance with expert systems are summarized, with examples, in Figure 1.

		System Performance Uncertainty	
		Low	High
Knowledge Base Specificity	Low	Internal Market e.g., spelling checker	Team e.g., auto repair
	High	Bureaucracy e.g., benefits advice	Clan e.g., C3I

Figure 1. Expert System Management Structures with System Examples

The framework has been subjected to limited tests with some promising results (Bills, 1987; Jones, 1987; Walker & Weber, 1984). Further work is required to more completely document the advantages and limitations of this approach.

Implications for System Administrators

The implication for system administrators is to identify dimensions of exchange tasks expert systems are required to support. Two dimensions central to the selection of an appropriate organization design are performance uncertainty and asset (knowledge base) specificity. The relative position along the exchange dimensions provides a guide to the

172

administrator for structuring their organization to efficiently and effectively achieve goals using expert systems.

References

Bakos, J. Y., & M. E. Treacy (1986). Information Technology and Corporate Strategy: A Research Perspective. MIS Quarterly, June, 107-119.

Ballou, D. P., & G. K. Tayi (1985). Reconciliation Process for Data Management in Distributed Environments. MIS Quarterly, June, 97-107.

Bills, D. B. (1987). Costs, Commitment, and Rewards: Factors Influencing the Design and Implementation of Internal Labor Markets. Administrative Science Quarterly, 32, 202-221.

Brathwaite, K. S. (1983). Resolution of Conflicts in Data Ownership and Sharing in a Corporate Environment. Database, Fall, 37-42.

Jones, G. R. (1983). Transaction Costs, Property Rights, and Organizational Culture: An Exchange Perspective. Administrative Science Quarterly, 28, 454-467.

Jones, G. R. (1987). Organization-Client Transactions and Organizational Governance Structures. Academy of Management Journal, 30(2), 197-218.

Kahn, B. K. (1983). Some Realities of Data Administration. Communications of the ACM, 26(10), October, 794-799.

King, J. L. (1983). Centralized versus Decentralized Computing: Organizational Considerations and Management Options. Computing Surveys, 15(4), December, 319-349.

Ouchi, W. B. (1979). A Conceptual Framework for the Design of Organizational Control Mechanisms. Management Science, 25(9), 833-848.

Ouchi, W. G. (1980). Efficient Boundaries. Working paper Graduate School of Management, UCLA, August.

Walker, G., & Weber, D. (1984). A Transaction Cost Approach to Make-or-Buy Decisions Administrative Science Quarterly, 29, 373-391.

Weick, K. E. (1969). The Social Psychology of Organizing Reading, MA: Addison-Wesley Publishing Co.

Williamson, O. E. (1981). The Economics of Organization: The Transaction Cost Approach. American Journal of Sociology, 87(3), 548-577.

Zmud, R. W. (1984). Design Alternatives for Organizing Information System Activities. MIS Quarterly, June, 79-93.

Some Propositions on Expert Systems for Military Managers

Major Tim O. Peterson
Logistic Management Center
Dr. David D. Van Fleet
Texas A&M University
Lieutenant Martha Ryan
9th Weather Squadron

Abstract

If knowledge is power, then managers who possess managerial knowledge have power over the organization's achievements. Expert systems purport to provide knowledge to their users. This paper presents a set of propositions about the usability and usefulness of expert systems in providing military managers with managerial knowledge.

Carroll and Gillen (1987) have recently identified a manager's knowledge base as being one of the key factors in a manager's ability to perform effectively. Kotter (1982a; 1982b) also indicated that a manager's knowledge base influences his/her ability to operate effectively. A manager's knowledge base consists of, at a minimum, the managerial skills, the manager's self-efficacy toward these skills, the relationships between the managerial skills and the management functions, and the linkages between these managerial functions and the organization's goals.

Three methods have been available to organizations to facilitate the acquisition of managerial knowledge in their new managers: 1) social learning, 2) training, and 3) education. Some organizations promote new managers into managerial positions without any training or education. New managers who find themselves in this situation "learn through experience" (Kast & Rosenzweig, 1985:419). The patterns of knowledge gathered from these experiences are often called heuristics or rules-of-thumb (Harmon & King, 1985). Heuristics are domain specific in that the pattern must match the current state of events to be activated. Over time and through exposure to many situations, a manager develops many of these patterns which (s)he stores in memory. When an event occurs, such as poor performance by a subordinate, the manager searches through these patterns looking for one that matches the current state of events. In this way, the problem domain is quickly reduced to a manageable size, assuming the manager can find a match between a cognitive pattern and the current situation. This type of knowledge has also been referred to as procedural knowledge (Greeno, 1973).

There are two risks in allowing new managers to develop their managerial skills in this manner. First, many new managers fail early in their careers because they do not have the necessary skills (procedural knowledge) to perform as a manager. Therefore, these new managers flounder when faced with managerial problems which they have not encountered before. Second, the lack of managerial skill (procedural knowledge) can also lead to managerial behaviors being constrained "because people hesitate to do things that they can't do well" (Kast & Rosenzweig, 1985:420). This is consistent with Bandura's (1977; 1982) self-efficacy findings. Often because a new manager cannot find an acceptable match between available rules-of-thumb and the current situation, the manager will avoid the situation. As Simon

(1987) points out, a common failure of managers is the postponement of difficult decisions because they cannot locate an acceptable pattern of actions and outcomes for the given situation.

One way to deal with this problem is to assign a mentor to new managers (Odiorne, 1985; Zey, 1984). A mentor is an experienced manager who teaches the new manager his/her set of heuristics for dealing with specific situations (Levinson, 1981). In this way, the new manager obtains a large set of successful patterns quickly. A new manager who has learned from a mentor behaves as if (s)he understands the underlying facts and relationships for the different managerial skills even though (s)he may not. Because a new manager who has learned from a mentor only has procedural knowledge, slight variations in the situation can generally cause him or her to reject a heuristic because it is not an ideal match with the present situation. To be able to use procedural knowledge over a wider range of situations, the new manager must also develop declarative knowledge (Greeno, 1973). Declarative knowledge is a broader understanding of the facts and the relationships between these facts from a specific domain. This type of knowledge is usually developed in training and education programs.

Organizations are keenly aware that managers need to develop declarative knowledge as well as procedural knowledge if they are going to perform effectively in their jobs. As a result, organizations will spend well over 20 billion dollars (Feuer, 1986; Gordon, 1986a; Stull, 1974) and will send over 9.3 million managers to some form of management development in 1987 (Gordon, 1986b). The sole purpose of these training programs is to improve declarative managerial knowledge and to increase managerial procedural skills. Unfortunately, it is unclear whether these training programs actually achieve their intended objectives (Bernardin, 1978; Ivancevich, 1979). Others (e.g. Jenkins, Reizenstein, Rodgers, 1984; Livingston, 1971; Leavitt, 1975; Maimon, 1980; Waters, 1980) have made similar criticisms of managerial education programs.

One reason for the ambiguity over the outcomes of both training and educational programs is that much of what a manager learns in such programs may not be used immediately upon return from the program. If the new skill is not used or practiced often, the organization runs the risk that any learning which did take place in the training program will decay and the knowledge will be "lost amidst the clutter in the manager's memory" (Norman, 1982:34). On the other hand, when a manager does combine declarative knowledge and procedural knowledge into organized chunks of information that are indexed and easily accessible, the manager is said to have compiled knowledge (Rector, Newton, & Marsden, 1985). In these cases, managers are very efficient in selecting the appropriate set of actions given a specific situation. To aid new managers in the acquisition, organization, storage, and retrieval of managerial knowledge, it seems that other methods must be developed which can provide managers with the appropriate knowledge and the necessary skills when they need them.

Expert Systems

Social learning, training, and education have all been used by organizations in an effort to provide inexperienced managers with their managerial skills. Each method has been problematic. Another potential solution to this problem may be the application of artificial intelligence to the acquisition of managerial knowledge. Knowledge-based expert systems are currently being developed in such managerial functions as planning (Duchessi, 1987), organizing (Blanning, 1987), directing

(Kearsley, 1987), and controlling (Slagle & Hamburger, 1987). More recently, knowledge-based expert systems have appeared which claim to provide managers with the necessary knowledge and skills to perform tasks such as performance appraisal (Kearsley, 1987), performance feedback (Blanning, 1987; Kearsley, 1987), employee selection (Blanning, 1987), and salary administration (Ellis, 1983).

While these claims seem promising, the question still remains whether these new computer-based systems can provide managers with the compiled knowledge to perform their jobs more effectively. As Dr. Arie Lewin (1986) said, "Whether these new systems help managers is still a question to be answered. We just don't know. At present, we are working on faith." Similiarly, Mr. Michael F. Rufflo, the cofounder of two companies developing managerial expert systems (Human Edge Software Corporation and AI Mentor, Incorporated), said:

> The technology is there, and so are many of the managerial models we need to create expert systems for managers. Those of us interested in the developmental aspects of expert systems are doing a good job of developing these systems. What we haven't done a good job of is determining whether these systems really make a difference. Of course, we believe they do. There's a lot of work to be done before we will have the answers, but from an organizational point of view, that's what's important (Rufflo, 1987).

While these empirical questions have yet to be answered, some propositions can be developed about the possible relationships which might hold. The general research question is: Do knowledge-based expert systems actually provide inexperienced managers with the skills they require to function like experienced managers? In this paper, the focus is on the usability and usefulness to the inexperienced manager. The expertise to build managerial expert systems has already been demonstrated, but it is pointless to build such systems "unless we take into account behavioral requirements on their usefulness and usability" (Carroll & McKendree, 1987:14). This lack of attention to usability has been identified as one key reason for the limited impact of expert system technology on some professions (Coombs & Alty, 1984; Kid & Cooper, 1985).

Usability

By using an expert system can an inexperienced manager more accurately identify the appropriate behaviors required in a specific managerial problem setting? One reason identified for inexperienced managers difficulties in these situations is their lack of managerial skills to perform these tasks (Maier, 1958). Expert systems purport to provide managers with the necessary skills (Blanning, 1987; Harmon & King, 1985; Silverman, 1987a). On the other hand, Rector, Newton, and Marsden (1985) report that studies where human experts have used an expert system have shown very little improvement in those human experts' decisions compared to human experts not using an expert system. They reason that, because the human expert already has compiled knowledge for this specific domain, the expert system contributes very little except to confirm the human expert's decision. Based on this information the following proposition is advanced:

Proposition 1: There will be an interaction between use of an expert system and manager's experience level on the accuracy of the specific managerial task to be performed. For inexperienced managers, use of an expert system and accuracy on the managerial task will be positively related. For the experienced managers, use of an

expert system and accuracy on the managerial task will be positively related but less so.

The next proposition is based on inexperienced managers using the expert system and inexperienced managers not using the expert system. Without the expert system, these two groups of managers should score equally on the managerial task considering that they are both groups of inexperienced managers. But, given that one of these groups of inexperienced managers is using an expert system which claims to provide the necessary knowledge to perform the specific managerial task, the following proposition is offered:

Proposition 2: Inexperienced managers using the expert system will perform significantly better on the managerial task than inexperienced managers not using the expert system.

Rector, Newton, and Marsden (1985) reported that human experts using an expert system show very little improvement when compared to human experts not using an expert system. These findings would indicate that there would be an insignificant difference in the scores on the performance feedback task between experienced manager using an expert system and experienced managers not using an expert system. Hence, the following proposition is posed:

Proposition 3: Experienced managers using the expert system will not perform significantly better on the managerial task than experienced managers not using the expert system.

If the findings of Rector, Newton, and Marsden (1985) are correct, little improvement in the managerial task should be observed from experienced managers using the expert systems, but inexperienced managers' scores should increase substantially. Since the scores of experienced managers using the expert system do not increase significantly from those scores of experienced managers not using the system, it is possible that the combined average score for the experienced managers would not be significantly different from the inexperienced managers using the expert system. In layman terms, the inexperienced managers with the aid of the expert system could perform as well as the experienced managers with or without the aid of an expert system. This lead to the following proposition:

Proposition 4: Inexperienced managers using the expert system will perform on the managerial task as well as experienced managers using the expert system and as well as experienced managers not using the expert system.

In addition to accuracy on the specific managerial task, a second measure of usability, self-efficacy, is important. One reason advanced for why managers resist certain managerial tasks is low self-efficacy expectations about their ability to successfully perform the task (Bandura, 1977; Bandura, 1982; Gist, 1987). On the other hand, expert systems purport to provide managers with the experience of many human experts in electronic form (Van Horn, 1986). These knowledge-based systems claim to provide managers with the necessary knowledge and skills to perform managerial tasks (Blanning, 1987; Kearsley, 1987; Silverman, 1987). If these system can provide managers with the necessary skills and knowledge to conduct the specific managerial task, then managers using an expert system will report higher levels of managerial self-efficacy. Based on this material, the following propositions are advanced:

Proposition 5: No significant interaction between the use of an expert system and manager's experience level on the managerial self-efficacy is expected.

Proposition 6: Use of an expert system will positively affect the level of managerial self-efficacy.

Proposition 7: Manager's experience level will affect the level of managerial self-efficacy. Specifically, experienced managers will report higher levels of managerial self-efficacy than inexperienced managers.

Usefulness

In addition to usability, it is also important to determine whether managers are satisfied with these systems. Usefulness of an expert system has been identified as an important empirical question which has not received adequate attention in the past (Carroll & McKendree, 1987). There are numerous studies (e.g. Epstein & King, 1982; Franz, Robey, & Koeblitz, 1986; Gallagher, 1974; King & Epstein, 1983; Miller & Doyle, 1987; Robey, 1979; Sanders & Courtney, 1985) which have reported the importance of perceived usefulness by the user in the acceptance of other computer-related systems. Goodhue (1986) posits that users perceive a system to be useful, that is they are satisfied with the system, when they see a fit between the task requirements and the functionality of the system. On the other hand, Rector, Newton, and Marsden's (1985) findings seem to indicate that perceived usefulness may be moderated by the experience level of the user. Therefore the following proposition is offered:

Proposition 8: Inexperienced managers using the expert system will report higher levels of satisfaction with the expert system than experienced managers using the expert system.

Another measure that has been used by some MIS researchers to measure usefulness has been to ask subject to place a dollar value on the system or its outputs (Gallagher, 1974; Zmud, 1978). Goodhue (1986) argues that this type of measure provides an overall perceived value of a system which can act as a surrogate for user satisfaction. Once again, the Rector, Newton, and Marsden's (1985) findings would seem to indicate that the value placed on the system may be moderated by the experience level of the user. Therefore the following proposition is advanced:

Proposition 9: Inexperienced managers using the expert system will report being willing to pay more for the expert system than experienced managers using the expert system.

Summary

Managers need knowledge and skills which expert systems may be able to provide. Based on existing research, however, the use of expert systems may not be completely straightforward. Nine propositions are advanced here to suggest the current state of our understanding of the application of expert system to management.

* References available from the first author upon request.

178

Complexities in Determining Optimal Work Group Size:
Task and Participation Effects

Glenn E. Littlepage, Ph.D.
W. Beryl West, Ed.D.
Debbie Beard, M.A.
Leslie Robeson, M.A.
Jeff Bryant, M.A.
Middle Tennessee State University

Abstract

Two studies were conducted to test predictions based on
Steiner's model of group productivity. In the first study, work
units of 1, 2, 5, and 10 persons worked on additive and
disjunctive tasks with differing coordination requirements.
Results supported Steiner's model for two of three tasks. A
second study using 2 and 10 person groups was conducted to
determine if participation patterns in large groups affect
performance. Results indicated that when participation is
required of all members, results support the predictions of
Steiner's model for all three tasks.

Research concerning the relationship between group size and group
performance has a long history, dating from Ringlemann's studies conducted in
the 1880's (Kravitz & Martin, 1986). Some studies of group performance show
that size does not affect performance (ie. Frank & Anderson, 1971), others show
relatively steady increases in performance with increasing group size (Ingham,
Levinger, Graves, & Peckham, 1974), still other studies show increased
performance but with diminishing returns (Yetton, & Bottger, 1983). Thus, it is
clear that group size may have an important impact on performance. However,
the nature of the functional relationship between size and performance seems
to vary with circumstances. Steiner's (1972) model of group productivity
provides a rich conceptualization of how the nature of the task moderates the
effect of group size on group performance.

Steiner has proposed that the actual productivity of a work group equals
the group's potential for productivity minus process losses. Both potential
for productivity and process losses often increase with group size. However,
the exact nature of these changes varies greatly depending upon the nature of
the group's task.

Potential for productivity is determined by the group's task relevant
resources (ie. knowledge and skills). Increasing group size may increase
potential, but the nature of this effect depends upon the task
characteristics, especially the way individual contributions can be combined
into a group product. Steiner (1972) describes several types of tasks, but
the present research focuses on two task types: additive and disjunctive. An
additive task is one in which members contributions can be summed to make a
group output. Tasks such as picking up litter or moving supplies are examples
of additive tasks. For additive tasks, potential increases with group size at
a constant rate. Disjunctive tasks are tasks which require the group to
select one member's contribution. Examples of disjunctive tasks involve
solving a puzzle or selecting among alternative solutions to a problem. For

disjunctive tasks, potential increases with group size, but with diminishing returns. On disjunctive tasks, new members add to potential only to the extent that they contribute non-redundant, task related resources. Since larger groups have more resources, each new member adds less to potential than previous members, and a point is reached when the group contains the required resources and new members add little if anything to potential.

Process losses reflect the extent to which group processes interfere with optimal utilization of resources to complete the task. Two categories of process losses have been identified: coordination losses and motivational losses. Both categories of process losses increase with group size, but the extent of losses depend upon the nature of the task. Tasks high in coordination demands should produce greater process losses as size is increased. Likewise, tasks that produce no individual performance measures should produce greater motivational losses with increasing group size (Latane', Williams, & Harkins, 1979). Steiner (1972) proposes that coordination losses increase with size at an accelerating rate. That is, the third person added to a group produces fewer coordination losses than the eighth person added to a group.

Increasing group size effects both potential for productivity and process losses. The nature of these effects depends upon the nature of the task. Thus, increasing group size should have differing effects upon performance of tasks with different characteristics. Two studies were conducted to test propositions of Steiner's model.

Study 1

This study examined performance of various size groups on tasks with differing characteristics. These tasks differed in coordination demands and in how individual contributions could be combined into a group product. Thus, for different tasks, performance should show different patterns of change with group size .

On an additive task with few coordination demands, Steiner's model predicts an almost linear increase in performance with increased size. A 33 item list of anagrams was chosen to represent an additive task. To keep coordination demands low, each group member had a list of all the anagrams.

On a disjunctive task with high coordination demands, Steiner's model predicts an inverted U function relating group size to performance. Initial increases in size should add needed resources and facilitate performance. After a point, new members would add little to potential but produce large process losses resulting in decreased productivity. A logic problem task was chosen to represent a disjunctive task with high coordination demands. Logic problems require logical deductions from clues to solve a problem. Since the group must select one solution, logic problems represent disjunctive tasks. Solution requires systematically following a line of reasoning and should be highly subject to distraction and require close coordination. To further increase coordination demands, clues were distributed among group members. Thus, for the logic problem task, performance should initially increase, then decrease as group size is increased.

A third task was used that involved a combination of additive and disjunctive elements and intermediate coordination demands. This task consisted of a 20 item quiz over various topics. Since the group must select one answer for each item, this task has disjunctive elements. But since individual items are independent and different members may answer different items, the quiz has additive (or divisible) components. Because the correct answer is not always obviously correct, selecting the correct answer requires coordination. Coordination demands were further increased by having members

share a single copy of the quiz and assign point values to each item depending
in their confidence in the group's answer. On this task potential should
increase less rapidly than on the anagram task and process losses should be
greater. The predictions from Steiner's model are less precise for this task
because of its combination of additive and disjunctive elements, but one would
not be surprised to find that increased group size produces increased
performance, but with diminishing returns.

Method.

A total of 324 undergraduate students were randomly assigned to work alone
or in groups of two, five, or ten. Each work unit worked on all three tasks.
Order of tasks was counterbalanced. To reduce problems on subjects in one
group communicating solutions to persons in later groups, three versions of
each type of task were used. These three sets of tasks were counterbalanced
with group size and order of presentation. Work units were allowed 5 minutes
to solve the anagram task, 25 minutes for the logic problem, and 15 minutes
for the quiz. Following each task, subjects completed a questionnaire
describing the participation in the group and their reaction to the task.

Results.

Scores for the anagram task consisted of the number of anagrams solved.
For the logic problem, scores indicate the number of correct deductions. For
the quiz, scores represent the sum of the (group assigned) point values of
items answered correctly. Results were analyzed separately for each type of
task using 4 x 3 ANOVAs: four work unit sizes crossed with three versions of
the task. Performance on the anagram task was higher in larger groups,
$F(3,60)=19.04$, $p<.001$. On the logic problem, performance differed across
group sizes, $F(3,60)=3.77$, $p<.02$. Performance on the quiz task was higher in
larger groups, $F(3,60)=10.35$, $p<.001$. See Table 1.

Table 1
Mean Performance Scores

	Group Size			
	1	2	5	10
Anagrams	1.39	2.50	3.83	6.17
Quiz	85.67	115.22	116.61	141.28
Logic Problem	3.22	2.94	4.44	5.28

Discussion.

Performance data for both the anagram and quiz tasks support Steiner's
theory. Performance on the anagram task shows an almost linear increase with
group size- the pattern predicted for an additive task with low coordination
demands. The quiz with its combination of additive and disjunctive elements
and moderate coordination requirements showed the predicted pattern of
diminishing returns. Performance on the quiz task increased with group size,
but at a negatively accelerating rate.

The logic problem failed to show the predicted pattern of performance
changes with group size. Steiner's model predicts that on a disjunctive task
with high coordination demands, performance will initially increase then
decrease with increases in group size. Data from the present study show an
increase in performance with group size, but not the predicted decline in

performance with the largest group size. Two explanations of the lack of a performance decline are plausible. One possibility is that group size was not increased to the point where the process losses of an additional person were greater than the potential derived from that person's additional resources. That is, perhaps performance would have declined if group size had been increased to 15 or 20 persons. This possibility is not implausible given the general nature of Steiner's predictions. Nevertheless, it seems that a 10 person group could be large enough to produce substantial process losses on a task requiring close coordination. A second explanation of the lack of a performance decline in the larger groups is related to participation patterns in groups.

Research on problem-solving groups (Bales,1951; Stephan & Mishler, 1952) indicate that as group size increases, the number of persons who participate very little increases. Thus, in the present study, increases in group size may have not produced corresponding increases in participation. If members did not participate in the group's discussion activities, it is unlikely that they produced additional coordination difficulties. Thus, it is possible that the failure to obtain the predicted performance decrements on the logic problem can be traced to a lack of participation by some members of the larger groups. Questionnaire responses provide some support for this interpretation: judgments of amount of information provided by others did not increase with group size.

Study 2

A second study was conducted to examine the impact of group participation on performance. In this study, 216 subjects were randomly assigned to 2 or 10 person groups. This study utilized the same methodology as the previous study with the exception of requirements to encourage all members to interact. Each subject was given a number of colored and white cards and instructed to deposit a card in a box each time he or she spoke. They were to start with the colored cards and when they were deposited, start depositing white cards. A subject who had deposited all his or her cards could speak only after all other group members had deposited their colored cards. On both the logic problem and the quiz tsks, each subject was given 5 colored and 10 white cards. Each subject was given 2 colored and 4 white cards on the shorter anagram task. This procedure was designed to insure some degree of participation by all group members.

Results.

Comparisons between study 1 which had no participation requirement and study II which required participation, indicated that required participation had little effect on performance of the 2 person groups on any of the 3 tasks. For the 10 person group, the interaction requirement did not affect performance on the anagram or quiz tasks. However, required participation reduced performance on the logic problem for 10 person groups, $F(1,30)=4.83$, $p<.05$. See Table 2.

Discussion.

The addition of the participation requirement affected performance in only one condition: the 10 person group working on the logic problem. This was the task with the highest coordination requirements. Without required participation, performance in this condition was higher than Steiner's model would predict. With the participation requirement, performance was more consistent with Steiner's model.

Table 2
Mean Performance Scores

	Study I		Study II	
	2	10	2	10
Anagram	2.50	6.17	2.33	6.83
Quiz	115.22	141.28	109.78	141.22
Logic Problem	2.94	5.28	3.33	3.67

General Discussion

Results of these studies are generally consistent with Steiner's model of group productivity. However, that model does not fully consider participation decreases in larger groups. Performance seems to be more closely tied to the number of persons actively participating in the group not merely the number present. A final implication is that the decreased participation of some members in larger groups is not always dysfunctional. On tasks with high coordination demands and with more than an optimal number of persons, decreased participation may result in improved performance.

Together these results indicate that it is difficult to determine the optimal group size. One must consider the characteristics of the task, and processes occurring within the group.

References

Bales, R. F., Strodtbeck, F. L., Mills, T. M. & Rosenborough, M. E. (1951). Channels of communication in small groups. American Sociological Review, 16: 416-468.

Frank, F. & Anderson, L. R. (1971). Effects of task and group size upon group productivity and member satisfaction. Sociometry, 34, 135-149.

Ingham, A. G., Levinger, G., Graves, J. & Peckham, V. (1974). The Ringlemann effect: Studies of group size and group performance. Journal of Experimental Social Psychology, 10, 371-384.

Kravitz, D. A. & Martin,B. (1986). Ringlemann rediscovered: The original article. Journal of Personality and Social Psychology, 50, 936-941.

Latane', B., Williams, K., & Harkins, S. (1979). Many hands make light the work: Causes and consequences of social loafing. Journal of Personality and Social Psychology, 37, 822-832.

Steiner, I. D. (1972). Group process and productivity. New York: Academic Press.

Stephan, F. F. & Mishler, E. G. (1952). The distribution of participation in small groups: An exponential approximation. American Sociological Review, 17, 598-608.

Yetton, P. & Bottger, P. (1983). The relationships among group size, member ability, social decision schemes, and performance. Organizational Behavior and Human Performance, 32, 145-159.

Colleagues: A Key to Instructional Effectiveness

David A. Carroll
U.S. Navy
Naval Amphibious School, Coronado

Abstract

Continuous controversy surrounds current practices utilized in the evaluation of instructor performances. Some critics have said the normal method of evalutation using students is of limited value when providing real growth-oriented feedback to the instructor. Since it seems that instructors want to grow and do their job better it is suggested a closer look be taken at using trained colleagues to provide the evaluations which will lead to better teaching and subsequently improved student learning.

Instructor performance is constantly being evaluated by students, staff, colleagues, peers and vicariously by organizations receiving the graduates. Beyond actual classroom performance, another standard practice is to write annual performance reviews and to give the periodic awards based on the outcomes of the instructor's performance evaluations. The focus of this study will be to examine the results of one instructor evaluation process. The results will strongly infer that the process is reliable and valid; thus being useful for instructor development and personnel evaluation.

Constant criticism of existing evaluation practices abounds in the literature (Medley, Soar, & Coker, 1983; Pittman, 1985; Singh, 1984). The subjectiveness of instructor evaluation was the point Medley focused on. He saw evaluation as open to bias and caprice and as such was essentialy no evaluation at all. Singh said ratings needed to be directed towards instructor improvement and not summative perfunctory grading systems. Therefore, evaluations which were to be useful for meaningful study and for feedback towards growth must objectively measure specific instructor behavior, mannerisms, and grammar (Medley et al, Pittman). Medley et al in further research found four specific steps were required to make the rating useful for growth: clarity of the task, objective evaluation criteria, qualified evaluation, and standard scores for comparison.

The greatest controversy in the literature has been concerned with who gives the most reliable and valid assessment of instructor performance: students or peers. (Centra, 1975; Medley, et al, 1983; Pittman, 1985; Singh, 1984) all agree that historically students have been the most reliable form of instructor performance. That is, students were consistent about who they liked and who they did not like. This level of evaluation was appropriately called "the clap" test. The "clap test" has it's place, but it is not an exhaustive measure of the quality of instruction received. Singh then went on to summarize that peer evaluation could be enormously effective if done properly and would add significantly to the level and quality of instructional evaluation. Centra demonstrated colleague ratings could be used to add validity to evaluations, an aspect often challenged in student ratings.

Based on the above findings and the state of training and development of the subject colleague evaluators, it was hypothesized that the evaluators would be a <u>reliable</u> and <u>valid</u> source of instructor evaluation. The null hypothesis of the study was "there will be no inter-evaluator correlation when grading the six instructors of the Chief Petty Officer (CPO) Leadership and Management Education and Training (LMET) course."

Method

Subjects

The study was conducted using personnel of a Navy Leadership and Management Education and Training (LMET) school as evaluators. These evaluators were in the LMET quality assurance branch and were concurrently LMET instructors and subject matter experts. All eight evaluators were pay grade E-5 - E-9 enlisted personnel and one officer, pay grade O-4. The evaluators had been teaching LMET for a minimum of two years each and some as long as seven years. Each evaluator had taught as a minimum 12-14 two-week LMET courses per year since being on staff. Also, each had been in the Navy at least six years with three in excess of 18 years. This study was done over a two-week period in February 1987.

The six instructors being evaluated were teaching the Navy's two-week CPO LMET course. Their pay grades were E-7 - E-9 and their experience varied from eight months to six years. All these instructors taught in two-person teams. Two of the six instructors were also evaluators. The other evaluators taught a similar but seperate two-week LMET course and were in a junior or in the officer's case a senior peer group to the instructors being evaluated. This close working relationship between the evalutators and instructors was what Singh (1984) considered ideal for development of a viable evaluation system. Each instructor was evaluated four times. To minimize peer bias, no two instructors were graded by the same group of four evaluators.

Materials

The behavioral evaluation form which has been standard with the Navy's LMET course since its inception in 1979 corresponded with the measures of instructor effectiveness used by Pittman (1985). The behavioral evaluation form has 32 factors combined into three groupings: creating learning conditions (CLC), managing group learning (MGL), and achieving learning objectives (ALO). Each evaluator had received a minimum of four weeks of formal training in the use of this form and had been using it constantly since first beginning to teach LMET.

Procedure

The procedural guidelines for using the behavioral evaluation form are in Figure 1. These procedural guidelines aligned with Medley's (1983) thoughts on developing high performing effective teachers. Also a subjective rating form was distributed to each of the evaluators. They were asked to complete it without discussion. The names were counterbalanced three ways on different forms to minimize a possible bias induced by a single order in the ranking.

The evaluators picked lessons spread across the three two-week CPO courses taught during this period of evaluation. All contracting and debriefing with instructors followed the steps in Figure 1.

Figure 1

Assessment Guidelines

1. Evaluation Review: Review Instructors past assessments.
   a. For trends improving or declining
   b. For areas of concern where changes should have occured.
   c. In case of continued ineffective performance:
      (1) Second occurrence: documented counseling session with action plan.
      (2) Third occurrence: outside assessment with video tape.
      (3) Ask for senior supervisor observation if no change effected.

2. Conduct contract brief with instructor.
   a. Set up time for feedback.

3. Observe lesson:
   a. Materials needed:
      (1) I.G. for lesson.
      (2) Student Journal.
      (3) Assessment Forms.

   b. Materials provided by instructor:
      (1) Notification of start time.
      (2) Road map and time lines.

4. Evaluation Review: Instructor's perception of lesson
   a. Items or sections done well.
   b. Items or sections not done so well.
   c. What I would do to improve parts not done well (action plan required).
   d. What I would do differently and reason for doing it differently.

5. Feedback: Evaluator's assessment of present lesson:
   a. Items done well (behaviors).
   b. Items not done so well.
   c. Items that need to change.
      (1) Not in accordance with Instructor Guide intent.
      (2) Not relevant or factual.
   d. Develop action plan with instructor for improvement.

6. Assessment's process review: instructor's perception of feedback session.
   a. What did you like?
   b. What did you dislike?
   c. How could the session have been more productive?

Results

The main thrust of this study was to empirically test the ability of these evaluators to give reliable and valid evaluations. The empirical test used was the nonparametric Kendall's Coefficient of Concordance "W" (Hendrick, 1981; Siegel, 1956). The subjective rankings of the six instructors are in Table 1. Kendall's Coefficient of Concordance "W" was 0.76. The significance using Kendall's corresponding F-test (Kerlinger, 1973; Siegel, 1956) was greater than 0.01.

Table 1

Subjective Rating Rating Results

Evaluator

		A	B	C	D	E	F	G	H
	1	1	1	2	2	2	3	3	2
	2	2	2	2	3	2	1	2	1
Instructor	3	3	6	6	6	5.5	6	6	4
	4	4	3	2	1	2	2	1	3
	5	5	4	4.5	4	4	4	4	6
	6	6	5	4.5	5	5.5	5	5	5

Kendall's Formulas: $\quad "W" = \dfrac{12 \ D2}{m2 \ (n) \ (n2 - 1)} \qquad\qquad F = \dfrac{(m-1) \ W}{1 - W}$

The objective test results using the behavioral evaluation form are recorded in Table 2.

Table 2

Objective Test Results

Instructor #1

Section	Evaluation totals				Totals
CLC	28	28	27	25	108
MGL	16	14	14	13	57
ALO	35	39	39	35	148
	x = 3.73				313

Instructor #4

Section	Evaluation totals				Totals
CLC	27	28	27	26	108
MGL	14	16	15	15	60
ALO	39	40	34	38	151
	x = 3.80				319

Instructor #2

Section	Evaluation totals				Totals
CLC	28	28	28	28	112
MGL	16	15	16	16	63
ALO	40	40	39	40	159
	x = 3.98				334

Instructor #5

Section	Evaluation totals				Totals
CLC	21	24	24	26	91
MGL	12	14	13	13	51
ALO	30	37	31	38	138
	x = 3.33				280

Instructor #3

Section	Evaluation totals				Totals
CLC	24	28	24	15	91
MGL	13	16	12	16	57
ALO	32	37	29	35	133
	x = 3.35				281

Instructor #6

Section	Evaluation totals				Totals
CLC	21	24	24	26	95
MGL	12	14	13	13	52
ALO	30	37	31	38	136
	x = 3.37				283

An overall "W" for this evaluation effort was computed using Table 3. This table lists the summary ranking of Table 1, the overall total ranking of Table 2, and the sectional (CLC, MGL, ALO) total rankings from Table 2. The overall "W" was 0.84. The association of all the rankings listed in Table 3 was also done using Kendall's Coefficient (Hendrick, 1981) in an identical manner as for Table 1. The procedures of Siegel, (1956) were used for checking the impact of ties. Ties were found not to be a significant factor.

Table 3

Composite Rankings of Subjective and Objective Test

Instructor	Table 1 Ranks	Table 2 Total Ranks	ALO	CLC	MGL
1	2.5	2.5	3	2.5	3.5
2	1	1	1	1	1
3	5.5	5	6	5.5	3.5
4	2.5	2.5	2	2.5	2
5	4	5	4	5.5	5.5
6	5.5	5	5	4	5.5

Discussion

These results imply a significant intercorrelation. Previous correlations doing similar colleague evaluations had results ranging from .10 to .57 with a mean in the .3 range (Centra, 1975). Here with an overall "W" of 0.84, significant at the 0.01 level, statistical correlation has been inferred and the null can be rejected.

The raters were statistically reliable over 84 indicators for six instructors. The premise of objectivity promoted by Centra, (1985) and Medley et al (1983) was reinforced. The overall group professional development proficiency evaluated by the average of the X scores of Table 3 was 3.59 on a 1 to 4.0 scale.

References

Centra, J.A. (1975). Colleagues as Raters of Classroom Instruction. Journal of Higher Education, 46, 327-337.

Hendrick, H.W. (1981). Behavioral Research and Analysis. Dubuque, IA: Kendall/Hunt Publishing Company.

Kerlinger, F.N. (1973). Foundations of Behavioral Research. New York: Holt, Rinehart and Winston, Inc.

Medley, D., Soar, R., and Coker, H. (1983). The Minimum Conditions for Valid Evaluation of Teacher Performance. Journal of Classroom Interaction, 19(1), 22-27.

Pittman, R.B. (1985). Perceived Instructional Effectiveness and Associated Teaching Dimensions. Journal of Experimental Education, 54, 34-39.

Siegel, S. (1956). Nonparametric Statistics for the Behavioral Sciences. New York: McGraw-Hill Book Company.

Singh, R. (1984). Peer-Evaluation: A Process That Could Enhance the Self-Esteem and Professional Growth of Teachers. Education, 105(1), 73-75.

Application of the Job Resource Framework to Flight Crew Management

Sharon C. Conley, Ph.D.
University of Arizona
Kathleen M. Conley
Captain, U.S. Air Force
Frank R. Wood, Lt. Col., Ph.D.
National Defense University

Abstract

This paper presents a framework for understanding how organizational resources affect the performance of the cockpit crew.

In recent years, aircraft accident investigation boards have increasingly recognized the contribution of human factors to aircraft incidents and accidents (Foushee, 1984). This heightened recognition has encouraged commercial airlines, NASA and the Air Force to search for better ways of managing aircraft flight crews. One such effort, "cockpit resource management" (CRM), concentrates on enhancing aircrews' abilities to use available (human) resources through training that emphasizes various crew coordination and communication techniques (Helmreich, 1986; Ginnett, 1987; Hackman, 1986).

CRM, while important, appears limited by two major factors. First, in concentrating on the dynamics of the cockpit crew team, the larger context of the parent organization (whether commercial, governmental or military) tends to be overlooked. Hackman (1986) offers this critique:

> cockpit crews always operate in an _organizational context,_ and the transactions between the crew and representatives of that context ... are consequential for any crew's performance ...we must look beyond our traditional focus on individual pilots...(to the) team- and organization-level (p. 23)

Second, CRM may overlook _other_ resources that are available (or unavailable) to crews, such as information, authority and materials. The identification of all resources that affect crew performance would appear of critical importance in both developing strategies to improve cockpit resource management and in considering the organizational context surrounding the flight crew. Much of the larger organization to which flight crews belong functions to provide resources that make them more effective on the job. To the extent that these resources are not available, crew effectiveness can be hindered. In this paper, we propose a job resource framework that addresses the resource flow between the larger organization and the crews. We offer some specific illustrations of how the framework could be useful in conceptualizing cockpit management within the larger organizational context.

The Job Resource Framework

The job resource framework was developed by Bacharach and Mitchell (1983), and is based on the assumption that employees in all jobs need sufficient resources to ensure effective job performance. Lack of resources contributes not only to poor task performance, but also to employee frustration and job dissatisfaction. The primary focus of the job resource framework requires identifying all of the resources employees need to perform their work and determining if these resources are present in sufficient quantity and quality.

The job resource framework identifies two key components that are present in any job: job resources and job resource dimensions. There are five primary job resources: authority, information, human support, monetary support and equipment. Summarizing Bacharach and Mitchell (1983) these resources may be described as follows:

Authority is the discretionary control of a responsibility or task. Does the employee have the ability to make the decisions required to complete the tasks for which he/she is held responsible or does authority (for particular decisions) reside elsewhere, e.g., with the employee's supervisor or other per-onnel? Information is what must be processed to make a decision in the face of uncertainty. Particularly in jobs with a high degree of uncertainty, (e.g., flying), information is a critical resource. In flying, information resources might range from immediate weather information to regulations and flight manuals. Does the employee have the basic information necessary to perform his or her job responsibilities? If not, where must he or she go to obtain the information and is it available when needed? Human support services are those human resources necessary for accomplishing job responsibilities. Coordination of human support would appear to be particularly important in occupations such as flying where work is conducted in teams (Ginnett, 1986; Hackman, 1986). Monetary support is defined as the budgeted and non-budgeted finances required for accomplishing job responsibilities. If the necessary financial support does not match the needs that arise during job performance, then efficiency and effectiveness are likely to decline. Finally, equipment is the physical resources employed in the process of performing job responsibilities. What equipment is needed to accomplish job responsibilities? Is the equipment used appropriately and is it of sufficient quality?

The second component of the framework is the job resource dimensions which specify factors affecting the flow of resources from the organization to the job. There are four dimensions: location, availability, timeliness, and quality. Location is the place in the formal organizational structure where a specific resource resides. Resources may be located relatively close to - or distant from - job incumbents. Availability is the accessibility of a resource to the job incumbent. This dimension assesses the ease in which resources are brought to bear upon tasks. Time-

liness is the match between the avail-
ability of the resource and the pace
of the work which requires it. Are
resources brought to bear on the task
at the appropriate moment? **Quality**
is the effectiveness of a resource
to meet the needs of the task. In-
formation that is timely, but of
poor quality, will hinder task
completion.

The job resources and resource
dimensions are summarized in Figure
1 with two additional summary
resources: time and energy. Even
if all of the resources are present
within an organization and accessible
to an employee, it is still possible
that the employee cannot accomplish
all of his/her job responsibilities
because of insufficient time and
energy.
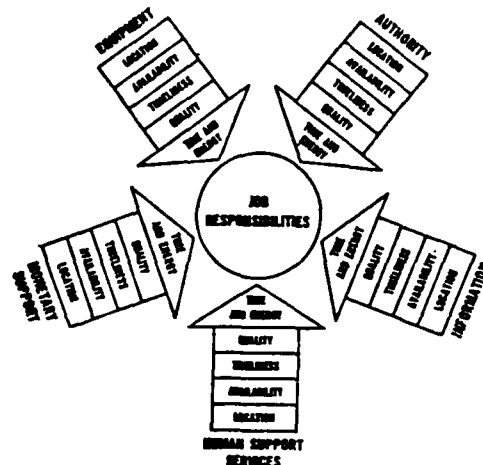
Figure 1. Basic elements
in the Job Resource
Framework (Bacharach and
Mitchell, 1983)

Application of the Job Resource Framework

We now explore some implications of job resources and resource
dimensions for cockpit management in the military context. In
providing illustrative examples, we primarily rely on resource
problems which, in our experience, are commonly encountered by Air
Force Military Airlift Command (MAC) crews.

Authority. MAC aircraft commanders, as well as commercial
airline captains, lack the authority to make many of the deci-
sions for mission completion. Instead, authority is located
elsewhere in the organization. Both must obtain permission from
Air Traffic Control before they take off or land (Hackman, 1986).
Unlike airline pilots, a MAC pilot who wants to depart more than
twenty minutes ahead of schedule to fly, for example, from Honolulu
to Guam must obtain permission from the Air Force command post at
Hickam. Thus, the **location** of authority for employees doing the
same job, but for different organizations, may be very different.
Further, the more distant location of this resource may affect its
availability and **timeliness**. Since the Hickam command post must, in
turn, coordinate with the Guam command post prior to approving the
early departure, problems with overseas autovon lines may mean that
the decision is not obtained in a timely manner. Recognizing this,
MAC aircraft commanders may decide against asking for early depar-
tures, even when weather may be more favorable for the earlier
takeoff.

Once resource problems are identified, this framework offers
a way of identifying possible solutions and considering potential
tradeoffs. For example, in the above situation, to increase the
timeliness of the authority resource, short of simply allocating

authority to the aircraft commander, the organization could upgrade the communications at overseas locations. Obviously, this latter solution would affect another resource, money, and the cost of this proposal would have to be weighed against other alternatives.

Information. On the C-141, the flight engineer is responsible for computing the airspeed to be flown on final approach as well as other significant data which the pilot requires for landing. To compute the proper data, the engineer requires information on weather at the destination. While pilots usually receive a weather briefing prior to takeoff, engineers must usually obtain updated weather prior to landing. The information they require is located at a ground meteorological station and may not be available due to radio limitations. Thus, updated weather might not be obtained until the aircraft is only minutes away from landing, adversely affecting the effectiveness of the flight crew. This is another timeliness problem. Since data takes approximately ten minutes to compute, the pilots may have to delay beginning their approach while the engineer finishes computing the data. Conversely, if the engineer elects to use the "old" data briefed to the pilots prior to takeoff, the problem becomes a quality problem as well. A possible solution to this dilemma lies in improving equipment. The Air Force recently purchased a fuel savings advisory computer which quickly calculates landing data once new weather information is obtained. Unfortunately, the present location of the computer in the cockpit places it out of reach of the engineer. This normally requires that another crew member load the weather update into the computer, decreasing the number of people monitoring the approach and looking out for other aircraft.

Human Support Services. The total number of individuals who contribute to a single aircraft launch is staggering, from the mission planner who coordinates the flight's itinerary to the maintenance personnel who perform the preflight. Even the driver of the crew bus performs a critical support function in transporting the crew to and from the aircraft and the base. In this regard, the bus and driver are normally located at motor pool; but might not be available. Further, even if they are available, they may not arrive in a timely manner. Lack of timely crew transport can delay a mission or cause a crew to feel rushed during pre-flight. Among the possible solutions are revising the base's priority system for allocating drivers or locating drivers and vehicles in closer proximity to the flight line. In fact, General Bill Creech turned around a declining sortie rate in the Tactical Air Command by fixing these kinds of resource problems (Peters and Austin, 1986).

Monetary Support. If an aircrew makes an emergency landing at a civilian airport, they will need funds for fuel and landing fees. Depending on the vendor, standard Air Force forms might not be acceptable. The money itself is located at the fuels management office at the home station. It is, therefore, available, but wiring $10,000 to the civilian aviation gas company may be a less than timely process. Further, the quality of the funds

may be a problem at a Japanese Air Base which is used to operating only in yen. Indeed, an aircraft commander once paid a fuel bill on his own credit card; one might wonder if that aircraft commander was pre-occupied with his finances (beating the bill home) the rest of the trip instead of approaches and air traffic control.

Equipment. To avoid other aircraft, pilots rely on ground-based radar equipment and flight control computers to augment traditional "see and avoid" procedures. The location of Air Traffic Control radar equipment is on the ground. So when Air Traffic Control computers break down, radar separation is no longer available and saturation of airspace around airline "hubs" may reduce traffic advisories. To take another example, in weather reconnaissance aircraft, poor quality radio equipment can cause flight crews to remain in dangerous storm centers (e.g., hurricane eyes) for up to forty minutes longer than necessary.

The importance of this model is that it provides a framework for understanding how organizational resources affect the performance of the cockpit crew. In the extreme case, it may explain how the demands of the flying task may exceed the capability of the aircrew, making them vulnerable for an accident. For example, we could argue that the organization has structured itself to provide sufficient resources to its crews for most routine situations. However, in some situations the cockpit crew can not effectively use these resources to accomplish their tasks, thereby reducing their effectiveness. Using this model, we would predict that the greatest potential for an accident lies in situations where the crew is operating out of a remote location, where needed resources are less available or not available in a timely manner, and where the quality of support might be reduced. These factors, combined with mission time constraints and energy limits (by the end of duty day or jet lag) may produce the greatest threat to aircrew performance. Even though the resources are provided by the organization, these factors may create a situation which demands performance exceeding the crew's capability. Granted, crews who are well managed (the CRM model) may fare better, but without the resources they need, they, too, may be at risk.

References

Bacharach, S. and Mitchell, S. (1983) Job Design Reference Manual, OAP, Inc., Ithaca, N.Y.

Foushee, H.C. (1984) Dyads and triads at 35,000 feet: factors affecting group process and aircrew performance. American Psychologist, 39, 886-893.

Helmreich, R.L. (1986) Theory underlying CRM training: psychological issues in flight crew performance and crew coordination. NASA Conference Publication 2455, 15-22.

Hackman, J.R. (1986) Group level issues in the design and training of cockpit crews. NASA Conference Publication 2455, 23-39.

Ginnett, R.C. (1987) First encounters of the close kind: the formation process of airline flight crews. Unpublished doctoral dissertation, Yale University.

Peters, J. and Austin, (1986) A Passion for Excellence, New York: Warner Books.

Group Composition and Behavior Supports as Determinants of
Aiding Attainment Behaviors and Group Performance

Nancy J. Stone, Ph.D.
Texas Tech University

Abstract

Groups, whose members' ability levels were either homogeneous
or heterogeneous, worked sets of crossword puzzles, which increased
in difficulty. The behavior supports facilitated either individual
or group behavior. The heterogeneous and group behavior support
conditions had significant additive effects on aiding attainment
behaviors. Aiding attainment behaviors were reliably correlated
with performance.

Group performance is a function of the surrounding environment as well as
the immediate conditions in which the group performs, intragroup
communication, and individual skill and intelligence (McGrath, 1984). Due to
the interdependence of these components, a human group is a system. Systems
operate along a continuum of stability. Hence, a change in any one part can
affect system stability. As system stability decreases, the necessary
behaviors which must be integrated and coordinated for a task to be completed
properly could include anything from communicating essential information to
actually performing part or all of a task which is not being attended to or
completed correctly by another. These behaviors will be referred to as aiding
attainment behaviors (AABs). The proper use of AABs ought to lead to an
effective and efficient group process. In order to understand the usefulness
of AABs, all components of the functioning group must be considered.

McGrath (1984) has identified various highly interdependent components of
group functioning. These include: task type, characteristics of individual
members, group structure, physical properties of the environment, and the
behavior setting. A short description of each component follows. Several
researchers have identified various task types (Lorge & Solomon, 1955;
Marquart, 1955; Steiner, 1966). Each type affects the manner in which the
task is approached, conducted, and completed.

Characteristics of individual members include one's desire for
maintaining group membership as well as one's ability to perform the task. A
person could maintain membership in order to seek prominence, to socialize, or
to complete the task (George, 1977). It has been shown that one's motivation
for maintaining group membership can be manipulated by properly designing the
operational environment (George, 1966; Kesterson, 1986; McDonald-Pierce,
1986). Members' ability will also affect group performance. If individuals
are selected and trained according to specified criteria for a particular
group, then each individual should perform at a highly competent level which
corresponds to the individual's training and ability level. These factors are
important because a group cannot function well unless its members are
competent (George, 1977). Furthermore, groups with no assigned leader and at
least one high ability member tend to perform better than such groups with no
high ability members (Laughlin & Bitz, 1975; Laughlin & Branch, 1972;
Laughlin, Branch, & Johnson, 1969; Laughlin & Johnson, 1966).

Group structure can be defined as the ratio between number of formal role
specialties and number of group members (George, 1977). A completely
structured group would have as many role specialties as members. In such a

case, each member's role would be distinct from all others. If no structure exists, then each member performs the same task. Group structure affects the social norms adopted by the group, the perceived status of each member, and the communication networks employed (Shaw, 1971). Therefore, group structure can influence whether information flows effectively to ensure that the task is completed in an efficient manner. This implies that group structure may inhibit or facilitate the implementation of AABs.

A group interacts in an environment which consists of physical as well as social properties (McGrath, 1984). The physical environment can strongly influence how the members behave (Barker, 1968). Behavior settings include the psychological and physiological effects and the physical properties of the environment, which exert various social forces that influence behavior. A behavior setting represents a certain fit between the structured group and the task demands, restrictions, and situations (McGrath, 1984). Now that the components of a group have ben identified, it is easier to understand the effects of AABs on performance.

Recall the notion of system stability. At a hypothetical level of complete stability, all components of the system (the functioning group) are at an optimal level. That is, the task is well learned, the members are competent to perform the task, and the physical environment is appropriately designed to help maximize group performance. As each component deviates from the ideal, e.g., the task becomes more difficult, then there is a decrease in system stability. George (1977) hypothesized that as system stability decreases, helping behaviors become facilitative. Decreases in system stability suggest that all components are not optimal. The greater the instability, AABs ought to facilitate performance and the correlation between AABs and performance should tend towards a significant positive value. When the system is stable, the group ought to be functioning smoothly and AABs tend to interrupt and debilitate that process. In this condition, the correlation between AABs and performance is hypothesized to be negative. There is an indifference area where AABs are neither facilitative nor debilitative. The level of stability is not so low as to warrant AABs, but they would not interfere with the group's performance.

The purpose of this study was to determine when AABs are most likely to occur and when they are facilitative.

Method

Subjects

Students (N = 120) were recruited from the undergraduate psychology pool at Texas Tech University. These students were volunteers and received bonus points for their participation. Participants were divided into 40 groups of 3 persons each.

Materials

The vocabulary portion of the Nelson-Denny Reading Test was used to determine each student's vocabulary level, as an indicator of intelligence, and as a measure of ability. The Worker Motivation Scale (WMS) was used to measure each individual's level of group motivation. To control system stability, the groups were required to complete crossword puzzles from three predetermined difficulty levels, easy, medium, and hard.

Procedures

This experiment was conducted in two sessions. During the first session, the vocabulary test and the WMS were administered. The information from the WMS was obtained for post hoc analyses. The vocabulary test was used to classify participants as high, medium, or low ability.

Once the individuals had been classified by ability level, they were contacted and scheduled for session two. A group's composition was either homogeneous (HOM), three medium ability members, or heterogeneous (HET), one low, one medium, and one high ability member. The groups were then assigned to a particular behavior support which either supported individual or group behavior. To encourage individual performance, each student was seated at an individual desk. The students sat around a single table to establish the group behavior support.

Each group had 4 20-min sessions to work on various crossword puzzles. Each 20-min session consisted of a predetermined set of 3 puzzles which were randomly assigned to the participants. The combined difficulty level of the puzzles increased over time which decreased system stability.

During all sessions, at least one observer, seated behind a one-way mirror, tallied the AABs of each member. The observers discriminated between active and passive AABs, according to the distinction described by Bales (1950). Giving information was active whereas seeking or asking for information was passive. The observers were trained in how to record the behaviors, and were blind to the groups' composition and to the objective of the experiment.

Results

A split-plot factorial (Kirk, 1982) analysis was performed. For this particular task, it was determined that quality was the best measure of performance. The percentage of boxes filled correctly out of the total number of boxes filled, represented quality.

HET groups displayed more AABs than HOM groups ($F(1, 36) = 2.91$, $p<.05$, one-tailed). Also, groups in the group behavior support (GRP) condition exhibited more AABs than groups in the individual support (IND) condition ($F(1, 36) = 4.09$, $p<.05$, one-tailed). The effects of group composition and behavior support are additive, i.e., there is no interaction; therefore, the GRP/HET condition represents the most favorable condition for producing AABs and the IND/HOM condition is the worst.

The recordings of AABs discriminated between active and passive behaviors. Active AABs were found to be significantly affected by the behavior support condition ($F(1, 36) = 6.04$, $p<.02$), but not quite by group composition ($F(1,36) = 3.61$, $p<.10$). There were no significant effects on passive AABs.

Quality was directly affected by group composition. During the last two sessions the HOM groups produced significantly lower quality than HET groups ($t_3(36) = 2.24$, $p<.05$; $t_4(36) = 2.52$, $p<.05$).

The correlation between AABs and quality was positive and reliable over all conditions and sessions ($r(38) = .35$, $p<.05$). Since group composition and behavior support have additive effects, the performance (quality) displayed by groups in the extreme conditions (GRP/HET and IND/HOM) was evaluated. The relationship between AABs and quality was reliable over all sessions for the GRP/HET condition ($r(8) = .90$, $p<.01$), but not for the IND/HOM condition ($r(8) = -.03$, n.s.). Further investigating the distinction between active and

passive AABs for the GRP/HET condition, the correlation between active AABs and quality was reliable over all sessions ($r(8) = .90$, $p<.01$), but the relationship with passive AABs was not ($r(8) = .46$, n.s.).

Finally, AABs were correlated with PM totals and TM totals (prominence motivation and team motivation scores obtained from the WMS). There was essentially no relationship between AABs and PM totals for the IND/HOM condition. Both active and passive AABs had a strong negative relationship with PM total for the GRP/HET condition at session 1 ($r(8) = -.69$, $p<.05$, for active; $r(8) = -.67$, $p<.05$, for passive), increasing to a zero relationship over time. The relationship between AABs and TM was unreliable for the GRP/HET and IND/HOM conditions over all sessions.

Discussion

To investigate the conditions under which AABs are most likely to occur and when they are facilitative, various components of group functioning were manipulated. A disjunctive task (a task which can only be completed according to the most competent member's ability) was presented to groups whose members' ability levels were either heterogeneous or homogeneous. The groups used in this study were leaderless and had no formal structure. Finally, the physical properties of the environment and the behavior setting either supported group or individual behavior.

The significant effects of group composition on quality over time indicates that the heterogeneous groups were capable of more nearly meeting the demands of the task as task difficulty increased, and the homogeneous groups were not. Since a disjunctive task was used, these results support the findings that groups with at least one high ability member will perform better than groups with lesser ability members. Therefore, group composition, which can be manipulated by selecting group members according to specific ability levels, tends to have an effect on a group's performance.

Group composition and behavior support have significant additive effects on number of AABs displayed. AABs, then, affect the quality of performance, depending on the group composition and the behavior support. Also, TM had no effect on AABs and PM was eliminated as the task became more difficult necessitating group interaction. The implications of these findings suggest that a person need not be concerned with personal differences, according to PM and TM, when making decisions on selection. Although previous studies have shown that groups whose members are high on TM perform better than groups whose members score low on TM (Kesterson, 1986; McDonald-Pierce, 1986), the process of selecting individuals who are competent as well as high on TM could be limiting. Rather, the emphasis should be on creating a behavior support which facilitates AABs and on establishing work groups with set competency levels to ensure that the majority of problems which may arise will be quickly and accurately resolved.

If it is not possible to set the behavior support to facilitate group interaction, then perhaps it is important to train the members in passive AABs in order to seek help as well as information about the situation. This ought to increase active AABs (giving information) which correlate reliably with performance.

References

Bales, R. F. (1950). Interaction process analysis: A method for the study of small groups. Cambridge, MA: Addison-Wesley.

Barker, R. G. (1968). <u>Ecological Psychology</u>. Stanford, CA: Stanford University Press.

George, C. E. (1966). Training for coordination within rifle squads. <u>Proceedings: U.S. Army Human Factors Research and Development Conference</u>, <u>12</u>, 80 - 86.

George, C. E. (1977). Testing for coordination in small units. <u>Proceedings: Military Testing Conference</u>, <u>19</u>, 487 - 497.

Kesterson, B. (1986). <u>The effect of worker motivation on group performance</u>. Unpublished doctoral dissertation, Texas Tech University, Lubbock, Texas.

Kirk, R. E. (1982). <u>Experimental Design: Procedures for the Behavioral Sciences</u>, 2nd Ed. Belmont, CA: Brooks/Cole.

Laughlin, P. R. & Bitz, D. S. (1975). Individual versus dyadic performance on a disjunctive task as a function of initial ability level. <u>Journal of Personality and Social Psychology</u>, <u>31</u>, 487 - 496.

Laughlin, P. R. & Branch, L. G. (1972). Individual versus tetradic performance on a complementary task as a function of initial ability level. <u>Organizational Behavior and Human Performance</u>, <u>8</u>, 201 - 216.

Laughlin, P. R., Branch, L. G., & Johnson, H. H. (1969). Individual versus triadic performance on a unidimensional complementary task as a function of initial ability level. <u>Journal of Personality and Social Psychology</u>, <u>12</u>, 144 - 150.

Laughlin, P. R. & Johnson, H. H. (1966). Group and individual performance on a complementary task as a function of initial ability level. <u>Journal of Experimental Social Psychology</u>, <u>2</u>, 407 - 414.

Lorge, I. & Solomon, H. (1955). Two models of group behavior in the solution of Eureka-type problems. <u>Psychometrika</u>, <u>20</u>, 139 - 148.

Marquart, D. I. (1955). Group problem solving. <u>Journal of Social Psychology</u>, <u>41</u>, 103 - 113.

McDonald-Pierce, L. G. (1986). <u>Motivation and productivity in small task-oriented groups</u>. Unpublished doctoral dissertation, Texas Tech University, Lubbock, Texas.

McGrath, J. E. (1984). <u>Groups, interaction and performance</u>. Englewood Cliffs, NJ: Prentice-Hall.

Shaw, M. E. (1971). <u>Group dynamics: The psychology of small group behavior</u>. New York: McGraw-Hill.

Steiner, I. D. (1966). Models for inferring relationships between group size and potential group productivity. <u>Behavioral Science</u>, <u>11</u>, 273 - 283.

The Study of Combat Behavior: Who Will Pursue the Beast in the Jungle?[1]

Lt Col John A. Ballard, Ph.D.
Air Force Institute of Technology

Abstract

Combat behavior has not been a major focus of behavioral re-
search. Reasons for studying combat behavior are given, followed
by discussions of why behavioral scientists outside the military
have conducted little research in this area and why the United
States military has conducted relatively little research.

Throughout history humankind has fought wars. This century has seen
millions die in combat. A look about the globe today finds hundreds of
thousands engaged in this most ancient of human activities.

Given the frequency and magnitude of this human behavior, one might think
that the corresponding psychological and sociological research would fill
countless volumes and libraries. This is not the case. While books by
military historians and prior combatants fill many shelves, sociologists and
psychologists (even military psychologists) have focused rarely on wartime
behavior and behavior in combat. To most behavioral scientists, combat is
analogous to Henry James' Beast in the Jungle--a lurking unknown presence with
which we live but about which we rarely speak--but when it is understood, may
provide profound knowledge, "knowledge under the breath of which the very
tears in his eyes seemed to freeze." (James, 1903/1981, p. 1171).

It is somewhat surprising that behavioral scientists associated with the
fields of military sociology and military psychology have not amassed a large
literature on combat in air, land and sea. Military sociology originated in
the classic work of Samuel Stouffer and his associates (see Stouffer, 1949)
who studied American soldiers in World War II (Coates and Pellegrin, 1965).
This work included extensive surveys of combat attitudes and experiences. But
military sociologists have not built systematically on the foundation laid by
Stouffer and his associates. Moreover, their focus has been on military
institutions in peacetime--not military institutions or social groups engaged
in war. Similarly, nonclinical military psychologists have functioned
primarily as industrial-personnel psychologists in military settings. Their
contributions to advances in selection and classification are legion. Their
contributions to understanding combat behavior are infrequent. On the other
hand, clinical military psychologists and clinical psychologists working with
former members of the military have developed a sizable literature on one
aspect of combat behavior, i.e., combat stress and associated post traumatic
stress disorders. But this is but one aspect of a much larger field one might
call combat behavior, or combat psychology.

---

[1] The views expressed herein are those of the author and do not necessarily
reflect the views of the Air University, the United States Air Force or the
Department of Defense.

This paper addresses the study of combat behavior by posing three questions and discussing answers thereto. First, why study behavior in combat? Second, why have behavioral scientists in universities and other nonmilitary research institutions conducted so little research on behavior in combat? Third, why has the United States military conducted so little research on behavior in combat?

Why Study Combat Behavior

Combat may be defined as "actual fighting engagement of military forces as distinguished from other military duties or periods of active service without fighting" (Webster's Third New International Dictionary, 1981, p. 452). As used here, combat behavior refers to individual and group behavior occurring in a combat environment--an environment in which hostilities have occurred, are occurring, or in which individuals perceive that hostilities are probable. With combat behavior so defined, domains for research cover the spectrum of conflict from reactions to terrorism and low intensity conflicts to high intensity conflicts and nuclear war.

Several reasons can be offered as to why one might expect behavioral scientists to conduct research on combat behavior. First, and most basic, is simply to advance our understanding of the human being. No portrait of the human condition would be complete if it excluded or failed to acknowledge combat behavior.

Second, the study of combat behavior provides unique opportunities "to look behind closed doors", to explore areas of human behavior not often seen-- or discussed. Combat brings to the surface "some of man's deepest fears: fear of wounds, fear of death, fear of putting into danger the lives of those for whose well-being one is responsible" and "some of man's most violent passions; hatred, rage, and the urge to kill" (Keegan, 1976, p. 16). In combat the behavioral scientist might see the limits of obedience and leadership, the nature of self-preservation, and "the far shores of courage" (Keegan, 1976, p. 33).

Third, it is difficult to imagine a human activity with a more profound impact on individuals, groups, and nations than combat. Throughout history combat has determined the fate of people, governments, and nations. Consider, for example, Agincourt, Waterloo, Bastogne. The significance of war in human affairs has attracted many scholars, very few of which have been sociologists and psychologists.

Fourth, the knowledge obtained by studying combat behavior has application to military forces. Creveld (1982) and others argue that effectiveness of a military force is a function of the quantity and quality of its hardware multiplied by its "fighting power", the social-psychological components of combat motivation. It is the human element that is often most decisive in combat (for further discussion, see Henderson, 1978, 1985). The potential of behavioral scientists to contribute to military force readiness through the study of combat behavior is considerable.

Combat Behavior and the Researcher Outside the Military

Given these reasons for the scientific investigation of combat behavior, why has there not emerged a research tradition for examination of combat behavior? Academe has independently generated very little inquiry into this area of human behavior. Research outside the military has typically been a direct response to a stated need by the military. Why have behavioral scientists, especially university based researchers, produced so little research?

Much of the answer can be found in the nature of research itself. Research always occurs in a social context. Questions to be studied are normally questions considered acceptable by the community of behavioral scientists, questions guided by the paradigm of the field (Kuhn, 1970). Similarly, research is never complete until it is embedded in a social context through publication and/or presentation (Runkel & McGrath, 1972). Social interaction thus shapes the scientist and the scientist's research.

Attitudes of behavioral scientists toward the military affect sociological/psychological research on subjects perceived as relating primarily to the military. In the United States, the behavioral sciences are "linked to the liberal tradition" and largely ignore study of the military (Janowitz & Little, 1965, p. 15). Glick (1972) commented that social scientists biased against the military tended to transfer that bias to fellow social scientists who study the military. He argued that "conventional wisdom to the contrary, studying the military . . . does not automatically make you a Dr. Strangelove" (p.7). Although academe may be more open to behavioral research on the military today than in the Vietnam era, any shift in attitudes is probably small. Furthermore, the number of social scientists directly familiar with the military through experience is likely to be comparatively small.

The behavioral scientist seeking to study combat behavior will find that combat behavior is not a recognized area of research within sociological and psychological paradigms. Apart from the literature on combat stress, the research is not highly integrated or widely recognized. Education in this area is not institutional. There are few extrinsic rewards that would encourage research. Moreover, the research itself would appear to be difficult. Experimental designs using introductory psychology students do not come readily to mind.

For some behavioral scientists, affective, emotional, or moral factors will always limit their interest in the study of combat behavior. The essential nature of combat has not changed. As Clausewitz stated, "blood is always its price, and slaughter its character" (1832/1968, p. 344). Even so, if ample funding were made available to support research on behavior in combat, research would probably be conducted.

Combat Behavior and the Military

It would seem that the United States military would have a vital interest in conducting and supporting research on combat behavior. Apart from numerous studies conducted in the late 1940s and early 1950s, no programmatic research dealing specifically with combat behavior is evident. Several reasons can be

offered for the lack of behavioral research on combat during the past thirty years.

First, there is not yet a general awareness in the United States military of the potential of the behavioral sciences to contribute directly to combat effectiveness. The status of the behavioral sciences in the U. S. military parallels the status of the behavioral sciences vis-à-vis the hard sciences (e.g., physics, chemistry). Behavioral science is often seen as "soft" or "touchy-feely". Furthermore, professional soldiers often view behavioral scientists "as naive" (Janowitz & Little, 1965, p. 15). Rampton and Innes (1985) state unequivocally that the Israeli Defense Forces (IDF) do not view behavioral science as a "frill" as is usually the case in the U. S. military. They report that "to create cohesive, well-led , and highly skilled fighting units, the IDF has made extensive use of behavioral science concepts borrowed largely from the North American scientific and applied literatures" (p. 220).

Second, the emphasis on technology and technological innovation can overshadow research on the human element. Some might argue that technological superiority is the key to victory on future battlefields. Henderson's (1978) analysis of the Vietcong underscores "fighting power" as a force multiplier in combat and the importance of research on the human element.

A third reason is perhaps the impact of quantitative management. The advent of large peacetime U. S. military forces after World War II brought an emphasis on managerial efficiency and associated management science disciplines, such as systems analysis and operations research. Segal (1985) argued "the quantitative management orientation has led defense planners to focus on factors that are relatively easy to measure, such as cost, and to overlook factors that are more difficult to measure, such as leadership, morale, and esprit de corps" (p. 204). That which is difficult to quantify may be more difficult to defend in budget processes.

Conclusion

Although combat has had a profound impact on individuals and nations, behavioral scientists, both in and outside the military, have conducted relatively little research on combat behavior per se. Knowledge obtained through research on combat behavior could potentially enhance combat capabilities, increase military readiness, and thus further peace through deterrence.

References

Clausewitz, C. von. (1968). On war (abridged version, A. Rapoport, Ed., of J. J. Graham translation originally published 1908). Harmondsworth, England: Penguin Books. (Original work published 1832)

Coates, C. H., & Pelligrin, R. J. (1965). Military sociology: A study of American military institutions and military life. University Park, MD: The Social Science Press.

Creveld, M. van. (1982). Fighting power: German and U. S. army performance, 1939-1945. Westport, CT: Greenwood Press.

Glick, E. B. (1971). Soldiers, scholars, and society: The social impact of the American military. Pacific Palisades, CA: Goodyear.

Henderson, W. D. (1979). Why the Vietcong fought: A study of motivation and control in a modern army in combat. Westport, CT: Greenwood Press.

Henderson, W. D. (1985). Cohesion: The human element in combat. Washington, DC: National Defense University Press.

James, H. (1981). The beast in the jungle. In S. Bradley, R. C. Beatty, E. H. Long, & G. Perkins (Eds.), The American tradition in literature (5th ed.). New York: Random House. (Original work published 1903)

Janowitz, M., & Little, R. (1965). Sociology and the military establishment (rev. ed.). New York: Russell Sage Foundation.

Keegan, J. (1976). The face of battle. Harmondsworth, England: Penguin Books.

Kuhn, T. S. (1970). The structure of scientific revolution (2d ed.). Chicago: University of Chicago Press.

Rampton, G. M., & Innes, L. (1985). The role of the behavioral scientist in support of future military operations. In J. G. Hunt & J. D. Blair (Eds.), Leadership on the future battlefield. Washington, DC: Pergamon-Brassey.

Runkel, P. J., & McGrath, J. E. (1972). Research on human behavior: A systematic guide to method. New York: Holt, Rinehart, and Winston.

Segal, D. R. (1985). Management, leadership, and the future battlefield. In J. G. Hunt & J. D. Blair (Eds.), Leadership on the future battlefield. Washington, DC: Pergamon-Brassey.

Stouffer, S. A., and others. (1949). The American soldier (Vols. 1 and 2). Princeton, NJ: Princeton University Press.

Webster's third new international dictionary. (1981). Chicago: Encyclopedia Britannica.

An Expanded Use for Psychological
Profiling in Combatting Low Intensity Warfare

Felix Moos
University of Kansas

W. Ronald Olin
Lawrence, Kansas Police Department

Abstract

The proven use of psychological profiling has been limited in
its application toward terrorists and others engaged in low
intensity warfare. Some of these limitations are the result
of ignoring the cultural values of the opposing forces. The
authors argue that an expanded concept in applying
psychological profile techniques could result in information
of value to the security of United States installations and,
perhaps, of significant value to law enforcement agencies in
this country in the immediate future.

The Changes in Warfare and Perceptions of the World.

Twenty years have elapsed since the 1968 Tet Offensive in Vietnam, a
battle that, in retrospect, may have changed our perceptions of warfare. It
marked a point in time where tactics using high technology were matched by a
different method and style of warfare--one to which Americans could not, or
would not, adjust. Since that time, low intensity, low cost warfare has
constituted the most immediate threat to free-world security.

Low intensity warfare achieves its goals incrementally. Its use is
based upon violence to intimidate the United States, and more importantly,
the United States presence and influence abroad. The world is literally
awash in low-level conflict. More than 40 countries, involving more than one
quarter of the world's nations, are experiencing some kind of conflict in
1988. Land, religion, ideology, and foreign occupation are only a few of
the multitude of causes that now drive these conflicts. Sun Tzu cautioned
that one must know the enemy and yourself or fear the results of your
ignorance. Unfortunately, we have not applied existing knowledge to under-
stand the tactics and the motivations of the practitioners of low intensity
conflict.

This paper is an attempt to redefine and suggests ways to broaden the
traditional psychological profiling techniques used by intelligence agencies
and law enforcement in order to apply them more accurately to potential
adversaries, insurgents, and terrorists who threaten the security of the
United States and United States installations overseas. The terrorist
challenge for the 1990s will require a far more integrated, multi-
disciplinary approach for productive and culturally specific threat analysis.

A more objective analysis of hostile groups and their leaders could be obtained by combining and modifying existing research techniques.

Psychological Profiling Techniques.

Profiling is an established psychological concept. As long ago as World War II, a top secret report by psychiatrist Walter Langer on the "mind" of Adolf Hitler was reportedly used by American and British leaders. Subsequent application of psychological profiling techniques by the FBI has been documented in articles in American law enforcement journals since a pilot project was initiated in 1978. Criminal profiling has been used by the FBI and other law enforcement agencies to provide two basic kinds of analysis: psycholinguistic analysis from written or oral sources, and psychologically based analyses of events based on the modus operandi of the criminal act. These analyses are based on the assumptions that one can obtain from crime scenes or communiques the psychological motivation of criminals. In many cases, these psycholinguistic or psychological profiles have been valuable tools in solving criminal activity. Considerable success has been reported by the FBI in its use of these techniques. Psychological profiling has had basic application in two important areas: criminal apprehension and hostage negotiation. One area in which profiling has been less successful is the understanding of terrorists' motivations and tactics.

One problem with the use of psychological profiling is that most of the research appears to have been done by a very few individuals and applied primarily in the evaluation of serial murderers or serial rapists. A broader, and presumably more successful, scope and experience has been achieved through the use of "interdisciplinary" teams. According to the literature on profiling, the definition of "interdisciplinary" teams involves law enforcement professionals working with psychiatrists and psychologists in their attempts to solve outstanding criminal cases. However, recognizing and understanding the importance of additional factors in profiling techniques may have a profound impact on the future success of the concept. Walter Langer in 1942 had a very strong advantage over other researchers in his study of Adolf Hitler: Langer was a product of the culture and possessed the national character of the German/Austrian population from which Hitler came to power. Modern American researchers seem to have overlooked this powerful additional source of information.

The variety of research reports available suggests that most current profiling techniques involve American criminal activities perpetrated by Americans and evaluated by American law enforcement officials who can "see" the criminal's point of view. This American cultural bias may have some significant negative consequences in efforts to extend profiling techniques in the evaluation of foreign criminal activity, or even in future "American" crime trends.

Some demographers have predicted that by the year 2000, the population of the country will have changed to 49-52% Caucasian, 16% Hispanic, 16% Black, and 16% other with 8-10% of this later category Asian. The more heterogeneous the United States population becomes, the less likely that those who create the profiles will be able to rely on the psychological

propensities of the "typical American criminal personality". If these demographic projections are correct, then there are powerful reasons to begin a comprehensive and flexible expansion of profiling techniques. One benefit of the expanded study would be a better understanding of terrorists and terrorism. This, in turn, could be invaluable in its application to the United States military.

The Value of a Cultural Approach to Profiling.

Terrorism and low intensity conflict represent a new form of warfare requiring new counterstrategies to meet this threat. Just as forms of conflict have evolved, so have societies and their own particular cultures. Our own society is becoming more culturally heterogeneous. Other cultures in the world, such as Iran, are becoming more culturally homogeneous as the result of specific and repressive governmental policies. In addition to the more conventional ingredients of psychological profiling, one must consider the need to include a far greater appreciation of "cultural" coding.

The identification of terrorists has been hampered by our ignorance of basic questions: With which group(s) are we dealing? What are their choice of tactics? What are their political goals? Are they violence prone? Will they sacrifice their lives? The profile of a potential adversary represents a distinct culture as well as a particular psychology. We maintain that culture, in every case and in whatever geographical or ethnic setting, determines what is real and what is ideal in an individual's behavior. Certain actions are nearly culturally specific. Thus, the values and the attitudes which make an individual fight, or die, for a cause are as "real" a code as more accepted psychological characteristics traditionally employed in standard profiling. An individual's, or group's, cultural values and attitudes compel congruent choices in actual behavior. This makes a German terrorist distinctly different in thought and behavior from a Japanese terrorist.

A Proposal for Study.

The United States and its culture remains the benchmark for the world's successes, and as a scapegoat for its failures. America and Americans are today perceived, perhaps as never before, with ambivalence and increasing hostility. A great deal of this hostility is related to the apparent ignorance of United States representatives who are incapable of speaking the language, recognizing the culture, or viewing the affairs of the host country through any but an exclusively "American" bias. Many non-Americans see the United States as a paradox: a country whose military, diplomatic, and economic prestige has been waning in a period when its cultural influence is still very strong.

The security of the United States military installations is linked, in large part, to the specific conditions created by the areas and people in which they are located. The local culture must be taken into account as part of the total evaluation of terrorist or low intensity warfare threat. As the actions of groups such as the New People's Army in the Philippines, or Shiite fundamentalists in Lebanon, continue to influence American presence

abroad. Localized projections profiling terrorists' activities could be of value in assessing future threats.

The use of this kind of information could result in the creation of "mini-manuals" of threat analyses and recommended hostage negotiation techniques for each United States installation overseas. The distinct characteristics demonstrated historically by terrorists or terrorist groups are cultural, as well as psychological in nature. Expanded understanding of these concepts would provide us with the psychological and cultural idio-syncrasies that we have, heretofore, ignored.

The development of intelligence information for the security of military installations, and the future use of psychological profiling in domestic crisis situations will require powerful new tools. One clear path is the use of an expanded concept of region specific, multi-disciplinary approaches to profiling adversaries. In the comprehension of threat analyses, Americans must sublimate their own psychological and cultural blinders and begin to see aggression and terrorism through the eyes and values of the enemy.

References

Ault, L.A., Jr. and Reese, J. T. (1980). A psychological assessment of crime profiling. FBI Law Enforcement Bulletin, March, 1-4.

Biggs, J. R. (1987). Defusing Hostage Situations. The Police Chief, May, 33-34.

Brooks, P. R., Devine, M. J., Green, T. J., Hart, B. L., and Moore, M. D. (1987). Serial murder: A criminal justice response. The Police Chief, May, 37-45.

Douglas, J. E. and Burgess, A. E. (1986). Criminal profiling: A viable investigative tool against violent crime. FBI Law Enforcement Bulletin, December, 9-13.

Hazelwood, R. R. and Burgess, A. W. (1987). An introduction to the serial rapist: Research by the FBI. FBI Law Enforcement Bulletin, September, 16-24.

Langer, W. (1972). The Mind of Adolph Hitler. New York: Basic Books, Inc.

Miron, M. S. and Douglas, J. E. (1979). Threat analysis: The psycholin-guistic approach. FBI Law Enforcement Bulletin, September 1-5.

Ressler, R. K. and Douglas, J. E. (1980). Offender profiles: A multi-disciplinary approach. FBI Law Enforcement Bulletin, September, 16-20.

Strentz, T. (1982). A terrorist organizational profile: A psychological role model. An unpublished manuscript from the SOARS Unit of the FBI Academy.

Human Endurance and the Modern Battlefield [1]

David H. Marlowe, Ph.D. and James A. Martin, Ph.D.
Walter Reed Army Institute of Research

Abstract

Combat performance is hindered by psychological
casualties--individuals who, in response to the stress
of battle, lose the capacity to function effectively.
As warfare evolved, the manifestations of psychiatric
breakdown have changed accordingly. The advent of
nuclear warfare has engendered concerns about the
ability of soldiers and officers to cope effectively
with this new type of conflict. Much information
is available about factors that predispose soldiers
to psychiatric breakdown in conventional battle
situations; however, the generalizability of this body
of knowledge to the modern battlefield is controversial.
This paper describes psychiatric breakdown, discusses
factors known to contribute to breakdown, highlights
how these might operate on the modern battlefield, and
explores ways of minimizing psychiatric casualties in
future conflicts.

In the Army, we normally prepare for the worst-case scenario. Current
planning focuses on the TRADOC Central battle scenario: "Airland Battle
2000". The threat in this war, according to translations of Soviet Ground
Force doctrine, involves continuous echeloned attack by conventional forces
supported by chemical weapons with regular use of tactical nuclear weapons.
These tactics are designed to have great lethality and wide-ranging violence,
and to psychologically shock and behaviorally immobilize our forces.

We do not have the ability or experience to develop accurate predictive
models for describing the nature or extent of psychological breakdown in this
kind of warfare. It seems reasonable to assume, based on history, the nature
of human groups and individual behavior, that those factors known to prevent,
or mediate psychiatric breakdown and performance decrements in the past will
serve the same function in the future. The degrees and ways in which the use
of nuclear weaponry on the battlefield will alter these factors, however,

---

[1] The views of the authors do not purport to reflect the position of the
Department of the Army or the Department of Defense, (para 4-3, AR 360-5).

208

remain unpredictable. With these terms of reference, the following is an examination of the kinds of disruptions and stresses that any significant military conflict will create for its participants.

Combat Psychiatric Casualties

Combat psychiatric casualties—classified over the decades as Shell Shock (World War I), War Neurosis (WW II), and more recently as Battle Shock and Combat Stress Reaction—encompass a variety of dysfunctions. Basically, combat psychiatric casualties are people who, in the course of combat, lose their ability to function effectively for varying periods of time. The symptom checklist for combat psychiatric breakdown covers an extraordinarily wide range of possible responses. The critical defining aspect of the syndrome is that the soldier is no longer able to perform his normal functions, engage the enemy, and ensure his own survival. He may be temporarily immobilized, emotionally defeated for minutes, and self-restored to duty rapidly thereafter. He may be impaired for hours and require therapeutic intervention by his friends and leaders. He may be dysfunctional for days and require the intervention of mental health professionals, or he may be lost forever and never restorable to duty.

The symptoms of psychiatric breakdown in combat may include apathy, depression, confusion, lassitude, and anxiety. The symptoms are often indicative of a severe anxiety reaction, such as hypoactivity, hyperactivity, hysterical reactions, apathetic immobility, conversion symptoms (such as hysterical blindness, mutism, and hysterical paralysis), and a range of other disordered behaviors. The combat psychiatric casualty is protean in its possible forms. The signs and symptoms are significantly governed by the ecology of combat and the values, attitudes, and beliefs of the social era in which combat takes place. In short, they are battle and situation specific.

In the early days of World War I, for instance, the manifestations of acute stress were popularly believed to be evidence of cowardice. As a result, physical disability was considered a culturally appropriate reason for removal from combat, as was a highly visible shocked and numbed state considered to have been induced by the physical shock of exploding ordnance upon the central nervous system, i.e., "Shell Shock." Later in the war the alternative diagnostic category of "War Neurosis" was widely used (based on the understanding that behavioral dysfunction reflected not "cowardice" but the overwhelming of the soldier's psychological defense). This marked an important change in etiological thinking, a switch from physical to psychological causes.

World War II was a psychologically more sophisticated era with a different combat ecology. It was marked by a pattern of maneuver and movement combined with episodic static warfare of varying intensity, punctuated by breakouts involving a significant number of intense meeting engagements between contending forces. These differences produced symptom sets that were distinct from those that characterized WW I. The legitimacy of fear and stress on the battlefield and the conceptual acceptance of their ability to degrade behavior and create dysfunction became basic aspects of U.S. Army doctrine. It was powerfully expressed in the often-cited phrase, "every man has his breaking point."

This war repeatedly demonstrated that levels of stress breakdown were directly related to the actual combat scenario, interacting with the cohesiveness, training, quality of leadership, and past experiences of the units involved. Extremely high levels of psychiatric casualties resulted when inexperienced troops with inadequate leadership met highly skilled, better armed, better led enemies. Extraordinarily high rates of stress casualties also occurred when well trained, well armed, and well led troops were subjected to massive rates of indirect fire, particularly when such rates had not been previously experienced or expected. Disease, hunger, thirst, and sleep deprivation also led to above average rates of neuropsychiatric casualties. Furthermore, situations characterized by extremely high levels of anticipatory anxiety, continuing for several days and not ameliorated by the actual performance of the tasks often had devastating consequences.

In all of this it is important to remember that the stresses of battle generate a wide range of psychological and behavioral reactions which are not normally diagnosed as psychiatric or combat stress casualties. For example, we do not count as psychiatric casualties soldiers whose wounding or death occurred as a result of the impaired performance and diminished alertness induced by stress. We tend to keep the psycho-behavioral and the psycho-traumatic separate, but they are not. Highly stressed soldiers may do dumb things and get killed. Altered levels of awareness may lead to soldiers being run over by their own tanks, engaging or fleeing from shadows, ceasing to operate with the reasonable level of "wise fear" that keeps soldiers alive and underlies good combat performance.

Factors That Contribute to Psychological Breakdown

A number of factors are known to contribute to psychological breakdown in battle. They include the fog of war, the chaos of the battlefield, disease, heat, cold, heavy volumes of indirect fire, ambiguity and fear of the unknown, and a sense of isolation on the part of the individual or small group. Marshall (1978) attributed the extreme and unforgiving "loneliness of combat" to the lack of information that often characterizes the conditions of soldiers and entire combat units. In the absence of information, the combination of rumor and fear enhances the threatening qualities of the environment. The miasma of the unknown may be dramatically compounded by loss of formal leadership, loss of contact with higher leadership echelons, and the loss of key informal leaders, generating panic and radical misperception of the world.

Combat intensity and lethality also play critical roles. Lack of nourishment increases risk, since soldiers in combat tend to stop eating as hunger is suppressed by fear and anxiety. Unless there are well-organized systems and leaders who ensure that soldiers are fed, they can easily become anorexic as World War II soldiers often did. Hunger is often exacerbated by dehydration as soldiers in combat also tend to stop drinking unless deliberately kept hydrated. Even in moderately intense, prolonged combat, the soldier's condition may deteriorate to the point that he is incapable of ordinary combat effort. While the proximate causes are physiological, the generating causes are psychological and emerge from the ecology of battle.

To all of these direct contributors to breakdown and decrement, we must add lessons from Korea, Vietnam, and the last two wars fought by the Israeli Defense Forces. Psychiatric breakdown is more likely among soldiers who do not perceive themselves as well trained and who have poor morale. Noncohesive units that are in interpersonal and organizational disarray and that cannot function as teams are prone to breakdown. All of these factors impinge upon the soldier and unit to create the stresses that lead to performance decrement.

Factors That Protect Against Psychiatric Breakdown

Factors that protect soldiers against breakdown and performance disruption include the cohesion of the unit, the morale of its members, and the degree to which soldiers are bonded by ties of confidence and trust to their leaders and to each other. Also critical are soldiers' confidence in their training and equipment; their perceived ability to control, master, and operate within the combat environment; and the belief that they have the skills and behaviors that will enable them to counter, outwit, and defeat the enemy.

Unit cohesiveness, morale, and sustainability culminate from a series of interlinked processes. They emerge from the training experienced by soldiers, their confidence in their trainers, the social ordering of the combat group, and the relationships among officers, NCOs, and soldiers. They are shaped by the behaviors of leaders and the mutual commitment of leaders and the led to the cultural rules ithin the organization. This "cohesion complex" includes trust, confidence, a sense of high professional competency, and sustained, clear information flow. The operational outcome of "military unit cohesion" is the ability of the group co maintain a sense of social integrity and technical mastery in combat.

Other Considerations

This paper has portrayed the range of factors that degrade and destroy performance on the battlefield, as well as some of the factors that maintain it. Yet we must ask another set of primary questions about the nuclear scenario: What are the destructive and supportive factors that are brought to the nuclear battlefield and are generated within it? What kinus of organizational and human relationships are required to survive within it? What kinds of training and what sets of attitudes and perceptions will affect the outcome? Factors that may be terribly modest in our calculations may exert vast leverage in a given scenario. For example, all our recent wars have demonstrated that one very important factor that maintains American soldiers in combat is their belief in the excellence of the medical system that supports them. We normally lose very few of those who were not killed instantly or dead on arrival. Most soldiers believe that if wounded they will be saved.

What are the potential consequences of the tactical nuclear scenarios that we envision when we know that our medical care system will not be able to deliver the effective response represented by the Vietnam era medivac? If the unit or squad is to maintain itself in a radically decentralized and violent nuclear combat scenario, we must provide it with the skills and training that will maintain disciplined combat behavior and prevent chaos. Every soldier

will need basic medical skills. Buddy aid will b͏ critical in order to maintain an appropriate level of group function. We already have units with this capability, for example Rangers. How many of the units that we will place in a nuclear scenario have the Rangers' density of highly trained medics?

We should remember that we never get the war we prepare for. If we are wise, we will deal with issues such as unit cohesion and realistic training (including medical care) that history has shown us have the potential for maintaining the functioning of soldiers and their units. We also need to train our junior leaders to recognize the signs of stress and provide them with the skills to effectively manage combat stress reactions. These are the issues that matter.

Summary and Conclusions

It should be apparent by now that many factors add to the stress experienced by soldiers on the battlefield. They live surrounded by the unknown. If they are unsure of what to do next or whether help is forthcoming, if they are not quite sure that they can help themselves, if they do not trust their weaponry, and if their morale is not reasonable, no one can say what will happen nor devise nomographs to predict it.

It is fruitless to try to project actual psychiatric casualty rates on the tactical nuclear battlefield. We can only be certain that, in a nuclear version of a high-intensity war, psychiatric casualties will be extremely high. We should focus therefore on what we can do to minimize and prevent combat psychiatric breakdown to the greatest extent possible. Those are the traditional approaches: organize soldiers and train them realistically; create highly cohesive units well bonded vertically and horizontally; teach them how to use time appropriately; give them cross-echelon training so that the young specialist can do the job of the platoon sergeant when the platoon sergeant is gone, so that there are others with leadership skills to hold troops together. Help soldiers develop self-esteem and enmesh them in systems marked by trust, confidence, and competence.

References

Marshall, S.. L. A. (1978). Men against fire. Gloucester, MA:
    Peter Smith. (Original work published in 1947).

Stouffer, S. A., Devinney, L. C., Star, S. A., & Williams, R. M. (1949).
    The American Soldier. Prinston: Prinston University Press.

Estimated Effects of Ionizing Radiation upon the Performance of Selected Combat Crews

Michael A. Dore and George H. Anno[*]
Pacific-Sierra Research Corporation
Los Angeles, California

Abstract

A general model is developed for characterizing the expected performance of four types of tactical army combat crews when the individual crewmembers function at degraded performance levels due to acute exposure to ionizing radiation. The model is also applicable to other situations that degrade individual crewmember performance. The results provide performance data for larger scale U.S. Army models that simulate battlefield conflicts where nuclear weapons might be employed. Performance-level data are generated as a function of dose and time after exposure for each crew type.

This study is the latest in a series sponsored by the Defense Nuclear Agency designed to quantify the effects of intermediate-level gamma and neutron doses on the combat effectiveness of tactical forces. While previous work [Anno and Wilson, 1984] characterized the expected performance of individual soldiers, the aim here was to extend that work to small army combat crew units. Specifically, the objective was to determine the collective performance of a crew as a function of the performance levels of each of the crewmembers. The types of crews considered are shown in Table 1,

Table 1--Selected Army Combat Crews

Crew Type	No. of Tasks	Crewmembers	No. of Tasks	Brief Description of Engagement Actions	Normal Time (s)
M109 Howitzer Gun Crew	31[a]	Chief of Section	6	Set-up, aim, load,	103.2
		Gunner	3	and fire 2 rounds.	
		Assistant Gunner	2		
		No. 1 Cannoneer	18		
Fire Direction Center (FDC)	26	Fire Direction Officer	8	Acquire, calculate,	93.4
		Computer	3	and transmit target	
		Horiz. Control Operator	7	aiming data to the	
		(Observer and Battery)[b]	(8)	battery.	
M60A3 Tank Crew	49[a]	Tank Commander	19	Site target, emerge	28.1
		Gunner	18	from cover, fire and	
		Loader	6	reload 2 rounds,	
		Driver	6	and resume cover.	
M901 ITV Tow Crew	24	Squad Leader	7	Move to site,	157.8
		Gunner	8	sequentially fire	
		Driver[c]	2	and guide both	
		Loader	7	missiles and reload.	

[a]Includes two recoils.
[b]Fixed times.
[c]Inferred--no data.

along with a brief description of the specific combat engagement action of each.

Performance is expressed as the ratio of the time required to complete a specified engagement with normally healthy crewmembers (the normal engagement time) to the time required when the performance level of one or more crewmembers is degraded (slower). A combat engagement consists of a number of tasks shared among the individual crewmembers that must be completed according to a predetermined sequence. The performance of each crewmember is expressed quantitatively in terms of the time taken to complete his assigned tasks.

Analysis of Empirical Data

A large number of field measurements of the time required by trained military personnel to perform the tasks comprising the engagements described in Table 1 was assembled [Moyer et al., 1984; and Moyer and Lam, 1987]. This data was statistically evaluated in order to provide the appropriate input to a combat simulation code. Of principle interest were:

- representative mean times for each task;
- the variances about the mean times for healthy crewmembers; and,
- the detailed chronological sequence and heirarchy of task executions to allow construction of task interaction diagrams.

The results were then analyzed to arrive at a distribution function for variances based on a best-fit of data relating variance and mean time. The form of the distribution function (lognormal) and mean variance relationship was then assumed to be applicable to degraded conditions since direct measurement of such data is not feasible. Also included in the measurement data were total elapsed times for the entire engagement. These data were used as a check on the validity of the simulation results.

Four-Dimensional Functional Fit

A simplified expression for overall crew performance as a function of the performance levels of the individual crewmembers was derived. Consider a crew consisting of four crewmembers, each responsible for a specific subset of the full set of tasks making up the engagement. Assuming that all tasks are performed sequentially, the total time for the mission will be simply the sum of the time taken by each crewmember. Let T_i be the sum of the normal (undegraded) times for the tasks of crewmember i, $i = 1, 2, \ldots, 4$, so that the total mission time is

$$T_{tot} = \sum_{i=1}^{4} T_i \; . \tag{1}$$

The performance of a crewmember is defined as $p_i = T_i/t_i$, where t_i is the longer time required to accomplish the same set of tasks when the crewmember is degraded. Thus, p_i is always between 0 and 1. Similarly, the overall crew performance is given by $P = T_{tot}/t_{tot}$, where $t_{tot} = \sum t_i$. It then follows that

214

$$P = \frac{1}{\sum \left[\frac{\alpha_i}{P_i} \right]} \quad (2)$$

where

$$\alpha_i = T_i/T_{tot} \quad . \quad (3)$$

Equation (2) shows that each individual crewmember performance is reciporcally weighted by his fraction of the total engagement time when all are undegraded. The assumption of a series of sequential tasks used to derive this expression is true only for the TOW-ITV crew. The other three crew types require the crewmembers to perform tasks in a series-parallel fashion, where at various times during the engagement two or more tasks are performed simultaneously. The derivation of an expression to accurately describe these crew types required more extensive analysis.

A computer code was developed to simulate in detail the interdependence of all tasks comprising the standard engagement for each crew type. This code, CREW-III [Dore 1986], was used with the task times and variances discussed above. A selected set of 40 combinations of degradation levels for the four crewmembers was choosen to span the four dimensional space. For each combination the CREW-III code was run in a Monte Carlo mode for 100 realizations of the total engagement to yield the mean or expected time for each of the 40 degradation states.

The results were remarkably well fit by a more generalized form of Eq. (2)

$$P = \frac{1}{\sum \left[\frac{\alpha_i}{P_i} \right]^{\beta_i}} \quad . \quad (4)$$

The β exponents allow four more degrees of freedom to account for interaction effects between crewmembers. Both the α's and β's were determined from a simplex algorithm which mimimized the standard error over all 40 points. The starting guesses for the α's were determined from Eq. (3), and the β's were initialized to 1.0; in all cases the results converged rapidly to a distinct minimum. Figure 1 shows a few representative slices

$* = 1.0 \quad Y = 0.8 \quad \diamond = 0.5 \quad + = 0.3$
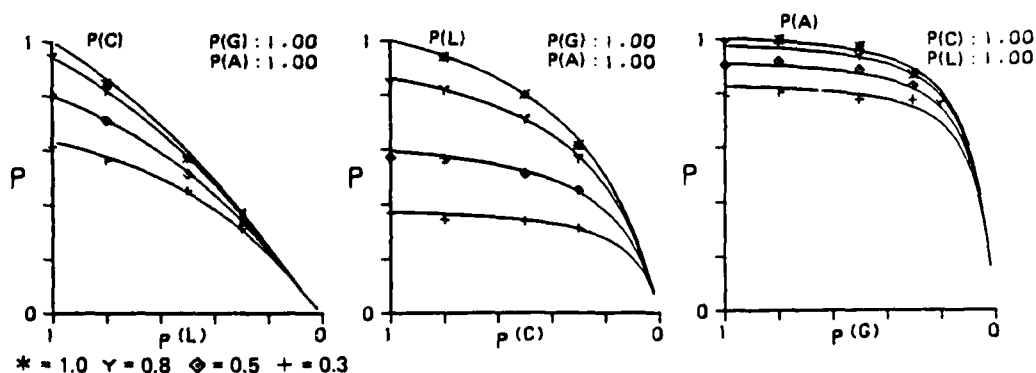
Fig. 1. M109 Gun Crew performance (P); Chief of Section (C); Gunner (G); Asst. Gunner (A); and Loader (L).

through the four-space for the Gun Crew, together with some of the data points to which this surface was fit.

Dose and Time Dependence

The above results are suitable for use in many applications, among them the Army Unit Resiliancy Analysis (AURA) code at BRL. AURA allows reconstitution of a tactical unit with any mix of degraded personnel, regardless of the underlying cause of the degradation. For instance, the limitations on crew performance caused by wearing MOPP gear in a chemical attack environment can be simulated by applying Eq. (4).

Other battlefield simulation codes are primarily concerned with effects on a tactical unit directly, rather than indirectly through the performance of the crewmembers. One such code is the JANUS conflict simulaton at LLNL. Here the purpose is to simulate the time dependent performance of a unit after acute exposure to an intermediate dose of ionizing radiation. Such a model extends naturally from earlier work [Anno et al., 1983] in which the performance of individuals was characterized for acute exposure as a function of dose and time. A given external dose to a unit (e.g., a tank) is reduced by an appropriate radiation shielding factor for each crew position, resulting in a dose to each crewmember. For a given time later the performance level of each crewmember is found from the data in Anno et al. [1983], and Eq. (4) is then applied to yield the overall unit performance at that time.

A complicating aspect to the above procedure is that not all individuals develop any or all of the disabling symptoms of acute radiation sickness which cause degraded performance. The crewmember performance results given in Anno et al. [1983] are conditionally dependent upon an individual having expressed the symptoms, and for those individuals who are not affected there is no degradation, i.e., their performance level is 1.0. This complication is accounted for by finding an expected performance for the unit by combining the sixteen possible degradation states of four crewmembers, each either expressing or not expressing symptoms, weighted by the incidence fractions for developing symptoms.

Estimates of incidence fractions for each of the major symptoms (e.g., nausea, vomiting, fatigability, etc.) of acute exposure to ionizing radiation have been determined as a function of dose [Anno et al., 1985]. While these incidence fractions are specific to each type of symptom, they are highly correlated and the actual performance degradation of an individual is usually dominated by the initial and most severe symptom developed. This allows an overall incidence curve to be determined by assuming an individual develops the characteristic degradation with the onset of any one symptom.

The expected performances were then calculated for the four crew types. Three dimensional and contour plots of performance for the Gun Crew versus log dose and log time are shown in Figs. 2 and 3.

References

Anno, G. H., and D. B. Wilson (1984). "Estimated Effects of Ionizing Radiation upon Military Task Performance: Individual Combat Crewmember Assessments," *Proceedings of the Ninth Symposium of Psychology in the Department of Defense*, 95-99.
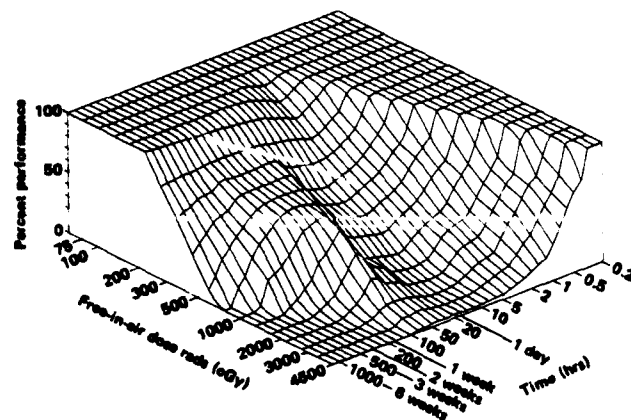
Fig. 2. Gun Crew performance surface.
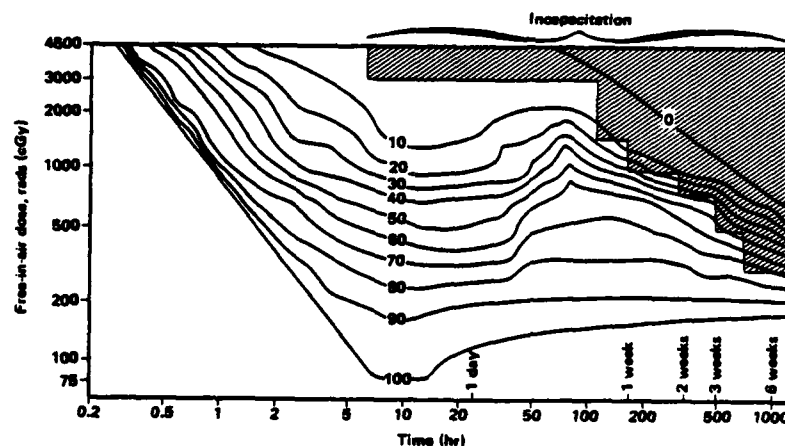
Fig. 3. Gun Crew performance contours, 10 percent performance
         intervals.

Moyer, M. J., T.K. O'Donoghue, and M.L. Feinberg (1984). *The Final Report of
the Task Timeline Analysis Project.* SAIC Corporation, McLean, Virginia, Report
No. 1-262-03-509-02.

Moyer, M.J. and S. Lam, SAIC Corporation. Unpublished. *M60A3 Combat Tank
Data.*   Forwarded in a letter dated March 6, 1987 to R.W. Young, Defense
Nuclear Agency, Alexandria, Virginia.

Dore, M.A. (1986). *CREW III User's Guide*, Pacific-Sierra Research Corpor-
ation, Los Angeles, California, Report 1660.

Anno, G. H., D. B. Wilson, and M. A. Dore (1983). *Acute Radiation Effects
on Individual Crewmember Performance*, Pacific-Sierra Research Corporation,
Note 572.

Anno, G. H., D. B. Wilson, and S.J. Baum (1985). *Severity Levels and Symp-
tom Complexes for Acute Radiation Sickness: Description and Quantification*,
Pacific-Sierra Research Corporation, Los Angeles, California, Report 1597.

An Evaluation of Stress, Confidence and Performance
Associated with Toxic Agent Training

LTC Frederick J. Manning, MS, USA
COL C. Fred Tyner, MC, USA
LTC Marvin A. Oleshansky, MC, USA
Walter Reed Army Institute of Research

Abstract

This evaluation measured the stress associated with a training exercise involving chemical warfare agents and tested whether the exercise changes the confidence, credibility or performance of those participating in the training. The on-site evaluation included over 100 subjects for all tests and over 1000 for some. An additional part of the study, addressing credibility, was conducted away from the training site during the same approximate time and involved 240 non-Chemical Corps subjects. Biomedical, questionnaire and behavioral measures of stress were not in strong agreement, but on the whole supported the conclusion that the exercise is mildly stressful. Questionnaire measures of confidence related to working on a chemically contaminated battlefield were clearly higher for subjects trained with toxic agent than for subjects without such training. Combat arms officers and NCOs were strongly in favor of toxic agent training, for their own units as well as the Chemical Corps, suggesting that the exercise will enhance the credibility of graduates.

This report describes an evaluation of training at the new Chemical Decontamination Training Facility (CDTF), USA Chemical School, Ft. McClellan, Alabama. During training exercises at the CDTF, soldiers are exposed to small amounts of chemical warfare (toxic) agents in a controlled environment while wearing standard issue (MOPP IV) protective ensemble. To evaluate the utility of these exercises with toxic agent, we investigated the stress, confidence and performance associated with the training.

The Chemical School provides instruction to Chemical Corps personnel ranging from junior enlisted through mid-career officers. During instruction which includes both classroom and practical exercises, students learn to detect chemical warfare agents, decontaminate personnel and equipment, recognize and treat symptoms of nerve agent poisoning, assess chemical battlefield scenarios and do a variety of other related tasks. In the belief that realism enhances training effectiveness, each course ends with an exercise using actual toxic agents in the CDTF.

The CDTF exercise occupies two successive mornings, each starting with a detailed safety briefing. On the first day, students rehearse agent detection and vehicle decontamination in MOPP IV in the open air; no agent is used. On the second day, students work inside but encounter several military vehicles contaminated with approximately 5-9 cc of either concentrated VX or GB. They attempt to identify the agent on each vehicle and perform the appropriate decontamination routine. Students know in advance neither the identity nor the amount of the toxic agent employed.

Our objectives addressed two broad concerns generated by the use of toxic agents. First, are Chemical School programs substantially more stressful for including training with actual agent? This concern was prompted by the main argument offered in support of the CDTF that realistically demanding exercises will better prepare students to perform well in combat. The amount of agent is small and the precautions numerous and

obvious; perhaps the training is all so safe that the students will not take it seriously. Second, regardless of how stressful the CDTF experience may be, the final goal of training is not to create stress per se, but to enhance the student's confidence that his training, equipment and doctrine will be an effective counter to the use of chemical agent weapons by some future enemy. An additional goal is to enhance the credibility of the graduates in the eyes of those they serve with and under after they leave the school. Implicit in these goals is the assumption that increased confidence and credibility will mean enhanced battlefield performance. This paper reports on our evaluation of four hypotheses:

A. Chemical School courses including the CDTF will be more stressful than courses not employing toxic agent.
B. Soldiers trained in the CDTF will be more confident than those not so trained.
C. Soldiers trained in the CDTF will be more credible outside the Chemical Corps than those not so trained.
D. Soldiers trained in the CDTF will perform better than those not so trained.

Method

Subjects

We drew subjects from four courses: Officer Basic (OB), Officer Advanced (OA), Junior Enlisted (AIT) and Advanced Non-Commissioned Officer (ANOC). Exams and end-of-course critiques were collected from all students as part of the Chemical School's standard operating procedure and provided our performance and confidence measures. All other measures were collected from volunteers solicited after a thirty minute group briefing on the nature of the study, several days prior to the CDTF exercise. A decision on sample size was made for each dependent variable based on: resources required and available; estimates of how data collection would either intrude on the School's operations or distort the phenomena we sought to study; a statistical calculation reflecting estimates of population mean and variance and of the minimum differences we thought meaningful. Sample sizes ranged from 120 (heart rate) to 1052 (confidence). For hypothesis C, we surveyed a group of 240 company level NCOs and officers outside the Chemical Corps, at sites in the U.S. and Germany.

Dependent variables

Stress was assessed through biomedical and psychological measures generally associated with "stress" responses. Biomedical measures included heart rate, blood pressure, blood hormones (cortisol, prolactin, ACTH and beta-endorphin) and overnight urinary hormones (cortisol and adrenalin). Heart rate was collected with a 4-lead battery powered portable recording system (Medilog) worn under the uniform by a subset of subjects from 0630 until the conclusion of the CDTF exercise each day. Psychological stress measures included: a mood adjective checklist which can be analyzed to provide indices of fear, anger, depression, fatigue, activity and happiness and which has proved sensitive in military exercises involving the chemical protective suit (1); group interviews of students and instructors participating in the exercise; student estimates of the risk entailed by the CDTF exercise; and direct observation of students during the exercise.

Confidence was evaluated with a questionnaire written especially for this study. It consisted of ten questions, to be answered on a seven point Likert scale, addressing the soldier's confidence in his ability and that of his classmates to survive a chemical attack, identify agent, decontaminate equipment, provide first-aid, and instruct others on how to do these things. It was given once, following completion of the CDTF exercise.

Performance was evaluated by examining the scores of written competence tests given by the School as part of its regular instruction process. Because all instruction and

testing is done before the CDTF drill, consistent changes in test scores might be attributable to an anticipatory effect associated with the facility. Although written School tests are certainly an imperfect and limited index of battlefield performance capacity, safety regulations and the collective nature of the decontamination exercise precluded any measure of hands-on performance in the CDTF itself.

Credibility was evaluated with a specially designed questionnaire given to NCO's and officers in combat arms assignments. Four questions assessed the extent to which respondents believed toxic agent training would have positive effects on their unit. One question asked the subjects how much he would pay, in unit training time, to get such training for his own soldiers.

Procedures

The Chemical School conducted, at our request and for the purpose of this evaluation, a CDTF "dry run" when the building first became available. The dry run involved all CDTF facilities and procedures but no toxic agent. Participants knew that no toxic agent was involved in these exercises. Because a major goal of the CDTF is to reduce the stress of a future chemical battlefield (by showing soldiers they can survive and function in the presence of a chemical warfare agent), the School also agreed to extend the CDTF exercise to three days for the eight classes from which we collected our stress measures. This allowed us to assess directly whether a successful decontamination exercise is associated with a decreased stress response to a subsequent exercise.

Hypotheses A,B and D were thus evaluated at the Chemical School through a one-time, unblinded comparison of measures taken before and after introduction of toxic agents to the CDTF. Major features of this procedure were a simple before-after contrast (dry run vs. wet run) with no crossovers and data collection starting three months before the CDTF opened and continued for three months afterward. Hypothesis C was studied away from the Chemical School with a short questionnaire at combat arms posts in the U.S. and Germany.

The data were grouped according to type of course (basic, i.e., AIT and Officer Basic; and advanced, i.e., Officer Advanced and Advanced NCO). The data for each hormone measured was subjected to a repeated measures analysis of variance (ANOVA) with agent (wet, dry) and experience (basic, advanced) as between-subjects variables. ANOVA was performed using the General Linear Model (GLM) procedure of SAS Institute's Statistical Analysis System.

Results

Stress

Overnight urine collection was undertaken to provide an integrated sampling method for cortisol and adrenalin released either in anticipation of or in response to the stress of the training exercises. Cortisol output was unaffected by toxic agent training, in either basic or advanced students, nor did it vary systematically over the three days of the exercise. Adrenalin output was higher in the toxic agent training periods (wet runs) than in the control training periods (dry runs), consistent with the expectation that working with actual nerve agent would be more stressful than an exercise with simulated agent. The difference in dry run and wet run adrenalin output was almost entirely confined to the basic course groups. Blood sampling just prior and just subsequent to the training exercise was designed to look at instantaneous stress hormone levels, at times much closer to the exercise itself than the overnight urine collection allowed. The best test of the hypothesis would have been samples taken during the exercise itself, a procedure precluded by subject safety considerations. None of the blood hormones

appeared to be affected by the use of agent in either the dry or wet runs, in either basic or advanced classes. <u>Blood pressure</u>, like blood hormones, were also sampled just prior to and just subsequent to each training session. Like the blood hormones, blood pressure was not affected by the training exercises. A similar conclusion can be drawn from our data on <u>heart rate</u>. Pulse rates were taken in conjunction with the blood pressure measures immediately prior to entering the CDTF training building, but unlike the hormone and blood pressure measures, heart rate was recorded throughout the exercise proper as well. The heart rate data analyzed were those collected during the first twenty minutes the subjects were in the training building. Heart rates during this time were markedly elevated over classroom rates. The increased heart rate can be reasonably attributed to anxiety rather than physical work since the subjects were by and large standing quietly, listening to instructions upon entering the building. However, both dry run and wet run subjects showed this elevation and ANOVA provided no justification for attributing this increase in heartrate to the use of toxic agent itself.

The biomedical measures of stress were supplemented by <u>psychological questionnaires</u> administered immediately prior to and subsequent to each day's training exercise. Subjects were instructed to answer the second of these questionnaires each day by describing their mood during their initial moments in the exercise itself. Scores on the "fear" scale of the mood adjective checklist were very low overall, averaging less than one on the 0 to 6 Likert scale, especially for the advanced course subjects. The scores nevertheless declined over the three day period, especially among the students working with toxic agent, who tended to have slightly higher scores before actually entering the training building.

Risk ratings for the CDTF exercises were assessed in relation to the subject's estimate of the risk of parachuting. Subjects were asked to provide these risk estimates on a 0 to 10 scale, where zero means "a completely no-risk activity" and 10 "the most risky or dangerous activity a person could possibly do". In sharp contrast to the phlematic reports of mood, the subjective risk scores associated with the CDTF exercise were quite robust. The ratings of the perceived risk of parachuting was 6.0 throughout the three day exercise. Students in the wet run did not differ from controls during the dry run in their assessment of the risk of parachuting. Prior to training, students in the wet run assigned the CDTF exercises a risk rating of 4.5, while controls rated the CDTF risk at 2.5. After three days of training at the facility, these ratings dropped to 2.5 and 1.6, respectively. Thus, students saw the CDTF exercise as considerably less dangerous after both toxic agent and simulant training.

<u>Direct observation</u> of student behavior in the CDTF revealed a striking difference between the dry run classes and their counterparts working with toxic agent. Students initiated departures from the training exercise more than six times as often when toxic agent was employed, strongly suggesting that training with toxic agent is more stressful than CDTF training without real agent. None of these students showed any clinical indication of exposure to agent, nor did any show the drop in red blood cell cholinesterase which would follow such exposure.

Confidence

Comparison of control and agent classes on each of the ten questions of the confidence questionnaire was conducted with the repeated measures ANOVA, using agent (wet, dry) and experience (basic, advanced) as between-subjects variables. Training with toxic agent had a highly significant effect. Subjects in the wet runs scored significantly higher than controls on every one of the ten questions. Experience also appeared to be a significant factor; basic course students generally expressed more confidence than advanced course students, although this difference reached statistically significant levels only on five questions. Hypothesis B, that soldiers trained in the CDTF with toxic agent will be more confident than those not so trained, seems solidly confirmed by these data.

221

Performance

Scores on the final exam for the chemical block for our subjects were the sole measure of performance employed. Even though in most cases this exam took place prior to the CDTF training exercise, it was hypothesized that anticipation of toxic agent training would have a positive effect on student motivation and be reflected in higher grades on the final exam. The data failed to support this hypothesis. Exam scores in general were quite high, averaging about 90%. Confining the analysis to those who scored less than 70, 80%, or 90%, did not alter the conclusion that the prospect of toxic agent training did not have an effect on final exam scores, i.e., just as many students scored below 70% in the agent classes as in the control classes.

Credibility

Scores on the seven-item credibility questionnaire were uniformly high with a modal score of 6 for every question. For each of the five questions, over 30% of the respondents gave this answer, expressing maximum approval of toxic agent training. Analysis of variance showed no significant effects of rank (NCO vs Officer), location (four bases), or specialty (eleven different fields).

Discussion

In summary, the biomedical, psychological and behavioral measures of stress described above provide modest support for hypothesis A, that CDTF training with toxic agent is more stressful than similar training not employing agent. The single most well-accepted hormonal measure of stress, adrenalin, was significantly higher in the classes using toxic agent, especially the junior enlisted and officer basic classes. Although few soldiers admitted to feeling much fear about the exercise, junior enlisted and officer basic students in the toxic agent training initially rated the CDTF exercise nearly as dangerous as parachuting. Further, soldiers found reasons to leave the training exercise six times more frequently when nerve agent was involved than when detection and decontamination were simulated.

Hypothesis B, that soldiers trained in the CDTF with toxic agent will be more confident than those not so trained, seems solidly confirmed by the consistently higher scores of the agent group on the confidence questionnaire.

The enthusiasm for training with toxic agent shown by the combat arms officers and NCOs who took the credibility questionnaire can be seen as a solid confirmation of hypothesis C, that soldiers trained in the CDTF will be more credible outside the Chemical Corps than those not so trained.

Final exam scores failed to provide any support for hypothesis D, that soldiers trained in the CDTF with toxic agent will perform better than those not so trained. A simple 'not proven' is probably the most appropriate verdict in this case, since the performance of most interest in setting up the program was not final exams, but battlefield execution.

References

1. Ryman, D.H., Biersner, R.J., and La Rocco, J.M. (1974). Reliabilities and validities of the mood questionnaire. Psychological Reports, 181, 497-484.

Impact of Stress on Combat Maintenance Organizations

Cheryl L. Batchelor
Air Force Human Resources Laboratory
Logistics and Human Factors Division
Wright-Patterson AFB OH 45433-6503

Abstract

The area of stress in combat has long been an issue in the military with specific research geared toward actual combat troops, such as the infantry and aircrews. Very little has been done in the area of the potential impact of stress on support troops; specifically, aircraft maintenance personnel. This paper provides a brief history of combat stress statistics, discusses the potential impact of combat stress in future conflicts on aircraft maintenance personnel, and finally, describes what steps the Air Force Human Resources Laboratory is taking to develop a program to reduce the impact of stress on aircraft maintenance personnel prior to the next conflict.

Statistics from specific battles during World War I indicate that anywhere from 14 to 60% of evacuees were due to psychological disorders. These numbers included personality disorders, neurotic and psychotic tendencies, and combat stress victims. During World War II the numbers were a little better in that only about 23% of all evacuees were specifically labeled "battlefield stress" casualties. The statistics continued to decrease where there was US military involvement. Combat stress accounted for only 6% of medically treated U.S. personnel in the Korean conflict and 3% in the Vietnam conflict.

With the continued decrease in the numbers of combat stress casualties through the years, one may question whether it is really necessary to pursue work in this area. Although not diagnosed as combat stress, there were many combat stress-type reactions which occured in the later conflicts; especially during Vietnam. In Vietnam, there was a much higher incidence of social problems, such as drug and alcohol abuse, sexual assault, disruptive behavior, and maladaptive behavior. For example, the Navy and Marine Corps experienced a 53% increase in the number of court martials from 1968 to 1969. Another major difference between Vietnam and Korea and other "wars" is that personnel were stationed in a "combat environment" for shorter, more definite periods of time.

Data from Israeli conflicts indicate that their combat stress casualties also decreased: from 30% in 1973 to 23% in 1982. Because of major differences in the Israeli and US life-styles, training, and battleground location, their combat stress statistics are considered optomistic in comparison with what US troops can expect in the future. Future combat stress statistics can be anticipated to be between 25 and 60% for US troops based on projected combat scenarios. One of the expected enemy strategies will be to attack from the rear so that no rear position will be available to US troops.

Much has been written about the effects of stress in our society, especially the stress upon executives, air traffic controllers, policemen, and others in demanding occupations. The military services are not immune to stress. Even in peacetime, many military occupations are a source of stress

because of high levels of responsibility or extreme physical and emotional demands.

The primary mission of an Air Force maintenance organization is to develop and sustain the skills required to produce sorties in a combat situation. Maintenance units must be prepared to perform this mission in any type of environment - peacetime or wartime conditions included. In the past, maintenance units have been protected, in a sense, by being located far enough behind the front-line to avoid major confrontations with the enemy.

The Air Force 2000 Planning Document and the Logistics Long Range Planning Guide predict our future combat aircraft maintenance units will not only be located near the front-lines, but they may also be working under very austere conditions. These working conditions will require accurate response, fast action, and clear thinking from personnel who have been stressed by a) rapid deployment; b) deployment with a group of individuals who are unfamiliar and uncohesive; c) having no realistic combat expectations; d) having no combat training of any sort; and e) disrupted eating and sleeping habits.

Combat stress, battle fatigue, combat fatigue, or whatever you choose to call it, has been examined from the military perspective in the areas of medical treatment units, aircrews and infantry troops. In addition, effects of chemical/biological warfare (including the use of chemical warfare protective ensembles) have also been researched from a stress perspective. Everyday life stresses, such as management stress, job stress, or family stress have also been investigated, but very little has been done with regard to the military support side: specifically, aircraft maintenance units. Those services with large numbers of combat fighting forces have indeed recognized the impact of stress during combat conditions and have taken steps to control and minimize that stress. The Army, for example, during World War I assigned mental health personnel to field units. Although that assignment concept was forgotten during the first part of World War II, it was reimplimented during the second part of the war, which increased the percentage of personnel returned to duty. The Army has plans to assign mental health personnel to field units again in future conflicts and plans to treat psychological casualties near the battle front. The Israeli Defense Force applied a modified version of the US Army's plan for forward field treatment of stress casualties during the 1982 war with improved return rates for combat stress casualties.

Academic literature is thorough in its documentation of the many debilitating effects of stress, and even in suggesting ways to limit or overcome stressful situations. The etiology of the stress to be encountered by aircraft maintenance units however, will be very unique from any other type of stress described thus far because maintenance personnel will be required to perform maintenance on the aircraft rather than defend themselves against the enemy. This scenario has not been experienced by the US Air Force or any other military organization; nor has there been training to any great extent on how to cope with stress. Therefore, it is unrealistic to expect an individual to perform up to the required standard without suffering some form of stress related degradation in quality and quantity of work performed.

The purpose of this paper is to raise the issue of combat stress, and to highlight its likely effect upon Air Force maintenance organizations. In addition, some ways of dealing with stress in order to minimize its long-term

effects upon the individual and the mission, and some possible R&D which could help minimize the occurence of stress will be discussed. The goal here is to highlight an issue that has not received adequate attention in the Air Force or the DOD as a whole. This paper will hopefully lead to the consideration of stress in war scenarios and in plans for sortie generation requirements.

Discussion

Combat stress can be expected to have a very significant impact on the ability of any unit in a combat environment to perform up to the required standards. Combat stress casualties from past conflicts came primarily from actual combat personnel. The question must be asked -- if trained combat troops can be expected to sustain such losses from stress alone, then what can non-combatant support troops, specifically aircraft maintenance personnel, be expected to sustain?

As stated earlier, in a combat environment, a maintenance organization will be required to prepare and repair aircraft to fly combat missions. This is a stressful task, even in peacetime because of the responsibility, time pressures, the necessity for quality maintenance, the long hours and extreme working conditions, organizational goals, and the real limitations of exper- ience and skill levels, across time demands. The projected combat environment will change things considerably. The peacetime stressors will still be pre- sent, but greatly intensified. Consider the following scenario, based on a Central European conflict.

The air base has been under frequent air attack with bombs, strafing, and chemicals. Ground attack from tank-led forces is threatening. Some aircraft have been moved to forward operating locations, but heavier maintenance must still be accomplished at the air base. Long hours, fear, injury, and death of buddies is commonplace. The fear of chemicals has become an added burden to bear. Personnel losses have severely limited the numbers and types of skills available, and damage has limited repair capabilities. The airplanes must still be repaired, refueled, and reloaded in such numbers as required to meet the sortie generation requirements. In the several days since the war began, the hyper-effeciency period suggested by Cowings, and reported by Kane (Tenth Psychology in DOD Symposium), is apparently ending. Fear, emotional fatigue, and physical exhaustion have begun to take their toll. These maintenance people have no real training on what to expect in combat, and they have no real Air Force role model or tradition of performing under combat conditions to live up to. The Wing Commander and the Deputy Commander for Maintenance are now faced with an additional loss of some percent for psychiatric casualties resulting from the stresses of the combat conditions -- combat stress. This percentage of psychological casualties in reality are losses not accounted for in the numerous manpower prediction models developed to date. These casualties are "physically and mentally unable to function -- although apparently suffering no organic damage." These are good maintenance people that the unit cannot afford lose, but they can no longer perform their duties. In addition to the impact of losing these personnel, consider for a moment a further impact of stress. Very difficult and technical tasks must still be accomplished by the remaining technicians who have also been stressed by the situation and must now pick up the slack of their lost "buddies." These tasks are often safety critical. The life of the pilot and survival of the aircraft may depentd upon correct performance.

Some very important questions arise at this point. First, at what point does fatigue, fear, and depression begin to impact the performance of a technical task? Second, is the stress induced by fatigue, and the resulting response, any different from stress induced by the combat environment?

Research in Combat Stress

Many studies have been done to study the impact of stress on aircrews. Specific areas that have been looked at are "fear of flight" and how stress impacts pilot performance. Dr. David Jones (USAFSAM/NGN) has done a vast majority of this work. Dr. Jones was a flight surgeon during the Viet Nam conflict and had first hand experience with the "fear of flight" syndrome. His expertise lies in the area of treatment rather than prevention. In his most recent technical report (USAFSAM TR 85-83), Dr Jones outlines the steps to be taken to set up appropriate treatment facilities and the care to be given to combat stress casualties.

The use of chemical warfare gear has been studied quite extensively in the areas of heat stress and claustrophobic reactions. To a much lesser extent, how the gear impacts performance in both quality and quantity have also been investigated, and continues to be researched as the chemical warfare gear is altered. There is currently a program being conducted at the Air Force Human Resources Laboratory to investigate ways to improve aircraft design and chemical suit design so the chemical suits will not be as susceptable to compromise. This effort is also looking at the additional time required to accomplish maintenance tasks in the chemical ensemble, and how practice in the chemical suit reduces that additional time required.

The performance of infantry and artillery troops has been under extensive investigation because of their front-line proximity and because of the type of jobs they are required to perform. Since World War II, there are some very specific statistics available on how stress impacts these military personnel and the development of various programs to control it. These statistics also include data from the 1973 and 1982 Israeli conflicts but again, primarily on combat troops. These programs include anything from general relaxation techniques to drug therapy.

Leadership and stress management programs are avaliable to the public and private sectors. These programs evolved from the 1960's when our society was going through a period of transition from simple and basic to the more complex society of today.

The combat stress reduction research program for maintenance personnel the Air Force Human Resources Laboratory is currently engaged in requires several assumptions be made based on past experience and literature. They are:

a) military personnel will suffer some type of reaction to stress in a combat environment. At some point, it will have an adverse effect on both quality and quantity of performance.

b) individual differences account for the varience in reactions to combat stress. These individual differences include personality type, demographics, and life-events.

c) a greater number of support personnel will be affected by combat stress than aircrew or infantry personnel. Because there are no role models available and no perceived impact on the conflict itself, their confusion will be that much greater.

d) because of the potential number of aircraft maintenance personnel affected by combat stress, the number of combat sories required will not be attained without some preparation or training for stress.

The study of combat stress in peacetime is very difficult. There are a large number of experiments which have been performed in the past which today could not be performed. Putting individuals into life-threatening situations and measuring their responses is no longer considered acceptable. Physiological measures have proven to be very effective but have one great flaw. How do you get the individual stressed to a level equivalent to what they might experience in a combat environment without causing unacceptable mental and emotional risk to the individual? This type of information would be required for a true validation study to be performed.

Based on the work of Dr. William Kane, a Summer Faculty Research Professor who was assigned to the Air Force Human Resources Laboratory in the summer of 1984, the following objectives were developed for the current program.

a) To understand the potential impact stress may have on aircraft maintenance personnel and how, in turn, that will impact their performance in a combat environment.

b) To reduce the negative impact of stress by developing programs which will raise the consciousness levels of aircraft maintenance personnel by increasing their awareness of future combat environments.

c) Build better individual coping skills for aircraft maintenance personnel so that they can continue to perform at their peak of quality and efficiency.

d) To develop some basic guidelines for senior NCOs and Officers to use in determining when stress has become detrimental to an individual's performance and what to do about it.

The first step to be taken is to see what research approach or approaches would be the most productive and feasible. Two different perspectives will be taken in determining what approaches should be more carefully evaluated. One approach will be from a hardcore maintenance and combat perspective. The other approach will be more from an academic perspective. The feasiblity of the recommended programs will be judged on the basis of such factors as cost effectiveness, the ability to be transitioned, and face validity. Once the feasibility of specific programs is determined, a follow-on program will look more specifically at developing those programs and field testing them.

References

Due to space limitations, a reference list is not included, however, a reference list can be obtained by contacting the author.

The Psychologist An Incarnation of Merlin?

Forrest Sherman, PsyD
Marine Security Guard Battalion
Quantico, VA

Abstract

The potential role of the psychologist as a modern day "shaman" is examined within the context of a military organization. Military organizations are examined as being heavily ritualized with a concomitant reduction in their ability to perceive reality in a flexible manner. The psychologist is perceived by the members of the organization as filling a "magical" role. This gives him the cultural permission to react to the environment in ways that are not limited by the traditions of that organization. A case is made that this role is an opportunity for psychologists to fill in the future.

Several years ago when having lunch with several aviators the trend of the conversation centered upon the lovely ladies they had met whilst traveling around the orient in their F-4s. As these aviators visited each area flight crews would spend the night working on the craft as the pilots and the NFOs (Naval Flight Officers) would "party." Noticing the hereldic devices the pilots wore on their flight suits I engaged in some fanciful specualtion. The pilots were clearly the Knights, the NFOs the Squires, the ground crews the Serfs, what role did I serve? I then remembered my flight helmet which my wife had decorated for me. It was covered with various symbols, the letter Psy, a group of trees (a pun on my first name), the astrological symbol for Aquarius, and a stylized rendition of my initials for a total effect reminiscent of a wizard's hat. I must be the wizard.

Shamans have played a ubiquitous role in societies. The role of the shaman has been studied in many cultures. Strictly speaking the term "shaman" applies to only certain cultural groups, primarily the arctic peoples with some cultural spreading to North American Indian groups and other limited groups in Europe. However, if some liberty of expression is allowed, the shaman plays a important role in many societies. He performs functions

228

related to healing, communication (especially via the spirit world), overseeing essential social ceremonies, appeasement of the spirits and advice. Generally the role of a shaman is most clearly found in hunting and gathering societies. These societies are often characterized by two characteristics, warfare and/or a dangerous environment. This is associated with an increase in the level of observed superstisious behavior. The shaman does not contribute to the actual economy of the culture and is often a burden to the culture, but his presence is widespread implying that his role is valuable to the survival of that culture.

Most of the functions that a shaman performs have little utilitarian value. His trancelike states and spirit communication do not change the patterns of the weather or of the herds of animals hunted. He often is incorrect in his predictions, and his interventions to heal illness often fail. However, cultures maintain the role of the shaman despite the lack of apparent value. His failures are ignored or explained away with remarkable ease. This suggests that the real role that he fills is not his apparent role but one less obvious.

The shaman is often "different" from the others in his culture. In true shamanism the mark of a shaman is that of some deformity, often extra fingers or teeth. This "difference" means that he is both a important member of the culture and at the same time uncharacteristic of the culture. This unique position allows him to operate with different rules and strategies then the culture specifies as normative. He is free to and encouraged to perceive the environment via different cognitive mechanisms. He may relate to others in ways that violate accepted norms. Nevertheless, his divergence does not lead to exclusion but he is placed in a position of respect and some authority. This serves to increases the total variance of the behavioral repertoire available to the culture. His being in a position of some respect within the culture validates the variance he contributes.

He is consulted and through that consultation permission is given for those within the mainstream of the culture to behave in unusual ways. His different perceptual strategies are used to broaden the perceptions of those who employ the normative strategies. It is this function that is important to the survival of the culture. The shaman provides diversity of perceptions and behaviors that would not be available to the society without his presence. A certain amount of balance must be achieved and his peripheral position allows for this. If he becomes too influential he will create too much change, if he is too peripheral he will not be available to the members of the society.

A literary example is seen in Merlin (of the King Arthur tales) who has been depicted in many variations with certain themes continuously appearing: He is central to the creation of King Arthur, his interventions leading to the conception of Arthur by sorcery, he is seen as Arthur's teacher, his advisor, his physician and lastly his prophet. Merlin was born without a father and was rumored to be the son of the "prince of darkness." Throughout the story of Arthur, Merlin plays the role of an active bystander. He does not conceive Arthur but only helps Uther in Uther's passion for Ygraine, he does not raise Arthur but only teaches him, he does not set the future but only tells of his visions over which he has no control. Despite his lack of active involvement, the entire Arthur legend requires the presence of Merlin.

Military organizations have some characteristics that can profit from the development of the shaman like role. Military organizations require a strongly defined world view. This is coupled with entrenched mores, perceptual strategies, and policies. Because of the extreme behavioral requirements of warfare, the leadership style and the prescribed interactions have become comparatively rigid and resistant to innovation. A classic example of this was the three year length of time the British General Staff continued to use frontal attacks during the WW I. This cultural resistance to change lead to the unwarranted deaths of hundreds of thousands. This rigidity should be balanced by a cultural mechinism that can introduce diversity.

Organization psychology has existed as a discipline and now the possibility is present for development of further roles within the military. If psychologists remember that they can provide a broadening of perception and problem solving strategies they will find that they can fill new and potentially helpful roles. If this comes to fulfillment, then we as a profession will have contributed even more to the mission of our respective organizations.

A psychologist can serve many roles, if he is part of an organization and not attached to a sepatate "medical" organization. Only by being percieved as part of the unit can he have any innovative solutions readily taken by the that unit. His very membership within the unit allows for his input to be more valued. The presence of a shaman role within the military culture serves the same function as it serves within a hunting-gathering society-diversity. As a Marine officer recently told me, "After working with you I am convinced that we need a psychologist at headquarters to keep us honest."

Developments on an Electronic Military
Leadership Bibliography[1]

David D. Van Fleet
Major Tim O. Peterson
Jon W. Beard
Texas A&M University

Abstract

This paper describes an electronic bibliography on military leadership available on floppy disks for IBM PCs and compatibles. The bibliography is an up-dated version of that contained in Van Fleet and Yukl's (1986) monograph. It has over 4,000 entries and can be quickly searched using the software provided.

In their monograph on military leadership, Van Fleet and Yukl (1986) include a bibliography with about 3,900 references. That bibliography was developed to be the single, most complete bibliography on military leadership available. Researchers and practitioners alike should find it valuable as a guide to ideas, examples, and sources of information. That bibliography has recently been up-dated and now contains over 4000 references.

Development

The initial development of that bibliography was broad-based. To provide integration and/or cross-fertilization, efforts were made to include all branches of the military by gathering references from military libraries which were asked to identify everything in their card catalogs under the headings, "leadership" and "military leadership". In addition to those references, other published bibliographies, technical reports, books, and articles provided material which was used in the bibliography. The huge body of literature on leadership in general was not included, however, unless it seemed to be used by those in the military as evidenced by its location in a source identified above.

Coding

Most of these bibliographic references are studies of or opinions about military leadership. Some are historical and/or biographical, some are business/industry related, and some cannot be classified. A few of the references are foreign. Finding references to subjects in which a person is interested is difficult since they are arranged alphabetically rather than by subject.

Coding would help. A tentative coding scheme for those references was presented at the last Psychology in the DOD conference (Van Fleet and Peterson, 1986). That scheme was specifically designed to aid military users, although it could be useful to non-military users as well. Using this coding scheme, a user could readily tell a great deal about the contents of each reference. By identifying references with the same code for a particular topic, users could locate all references of use to them. While adding this coding information to each reference would be a considerable improvement, it would still be a laborious task for users to sort through the 4,000 entries. It would not be, however, if the bibliography were available in an electronic format. In an electronic format, users could search rapidly for any one or even combinations of the coded references. This would greatly speed up searches but also enable users to extend their work since the effort to locate references would be so much smaller.

Electronic Bulletin Board

The most useful way in which to make this bibliography available in electronic format would be as a bulletin board system (BBS). In this format, users anywhere in the world could, via a modem, telephone the BBS, search by one or more codes, and see the references in which they are interested. Such a system could also enable them to obtain printed copies of the materials and to inform the operators of the system about additional materials which could be added to keep the bibliography up-to-date and complete. That system continues under development using a database system to further extend its usability (Van Fleet & Peterson, 1986). Such a system does involve time, software, hardware, and the availability of a telephone line. All of these necessitate money. The lack of funding to develop the system even on a trial basis has slowed progress in developing the system to a snail's pace, unfortunately.

Hypertext

Yet another way to make the bibliography more useful could involve the use of a hypertext system. Hypertext is a word which was coined some 20 years ago by Nelson (1965). Hypertext is defined as "a body of written or pictorial material interconnected in such a complex way that it could not conveniently be presented or represented on paper" (Nelson, 1965, p.96). The hypertext concept is attributed to Vannevar Bush's (1945) article, "As We May Think". In this article, Bush describes a machine or device called MEMEX (memory extender). This proposed device was an attempt to provide a convenient way for people to "manipulate' and "extract' knowledge from the "growing mountain of research" in which they are being "bogged down" (p.101). Bush continued with the observation that, "professionally our methods of transmitting and reviewing the results of research are generations old and by now are totally inadequate for the purpose" (p.102). Hypertext was seen as the resolution of this problem.

Present day databases (e.g. BPI, SSCI, Infotrac, etc.) of information do not operate in hypertext fashion. In current databases, information is placed in storage and then accessed by a limited number of key words of some form. The key words are assigned by library and information science techniques. Though helpful classification systems, the key word identifiers are sometimes of limited usefulness in locating information. If an exact match of the key

word is not found, a match is not made even if the information is really present. Bush (1945) noted that "the human mind does not work that way" (p.104). Human thought operates by association; "with one item in its grasp, it snaps instantly to the next that is suggested by the association of thoughts ..." (p.104). MEMEX, or a similar type of future device or technique such as the one described below, could use association and could be "a sort of mechanized private file and library" (p.105).

The military leadership bibliography of Van Fleet and Yukl (1986) has been placed in a new software system, FYI 300 Plus (hereafter referred to as FYI; Kimmel, 1986). This software can be used in hypertext-like fashion to facilitate the manipulation of the leadership database. FYI allows a database to be indexed on any word or phrase that exists within the system. This can include author's names, dates, and most any word in the title (single letters and some extremely common words, such as and, the, of,... are omitted from the indexing process through use of an "Omit List"). Key words can be defined and indexed, as in the large public access databases. Even words in abstracts, or entire articles, can be indexed if desired.

FYI allows searches of a database to occur using individual words or phrases, or combinations of words or phrases, linked together with logical operators. Though similar to the large databases, FYI can search for and match any word (subject to the omissions mentioned above) that occurs within the database. Exact matches are not necessary since searches can also be made using only truncated portions portions of words or phrases. It is then possible to view the location(s) at which the word occurs; in this case, to examine references on military leadership studies.

Though not a full hypertext system, FYI does allow the user to search for information along individualized paths using words or phrases that were selected by the user, not defined by someone else. FYI does, however, represent only a small portion of the potential of a true hypertext system. But, in true hypertext fashion, FYI does enable those interested to rapidly, accurately, and conveniently scan the over 4000 references in the military leadership bibliography.

FYI permits cross-indexing every word in the entry (except words in the "Omit List"). Since "full-text" cross-indexing is very easy, it has been done in this instance to make searching the bibliography as easy as possible. Thus, a user can select any word or combination of words to search the bibliography rather than simply a coding list developed by someone else.

After FYI retrieves the information you requested, it will either print the information, or send it to a disk file. When information is sent to a disk file, the new file is directly compatible with your word processing program. Thus, you can quickly include information retrieved from the bibliography system in a document that you are writing with your word processor.

FYI permits full boolean logic to be applied to searches so that you can search and retrieve what you want using AND, OR, XOR, and NOT search logic. This will result in a significant increase in your efficiency of handling the bibliographic materials. FYI also lets the user save standard search requests on disk for repeated use. Thus, if a user needs to do the same search

frequently, it is not necessary to re-enter the search request each time. It can simply be saved on disk, and called when needed.

An Example

As an example, consider a user who wanted to locate all of the references in the bibliography which involve the Air Force in combat. Combat could be directly searched for, but it could also be found under the word, war. However, if war were used, all of the war college references would also be located, so the user would want to exclude references to the Air War College at Maxwell AFB. This means sorting through 4,101 references and 74,592 words (although, on the system, only 9,469 key words are indexed). The search specifications for that user are (((AIR) \AND (FORCE)) \AND ((COMBAT) \OR (WAR)) \AND\NOT (MAXWELL)).

The number of references with the word "air" in them is 371; the number with "force" in them is 246; and the number with both "air" and "force" is 224. The number of references with combat in them is 110; those with war number 269; and those with either word total 378. However, the user's careful specification of the search pattern reduces the number of references markedly. That search takes approximately 10 seconds and yields the following 7 references (there would have been 31 references if the "\AND\NOT (MAXWELL)" limit had not been used).

Gates, E. "New look at the Air War College." Air Force Magazine, 1977, 60(1), 54-57.

Masters, R.E. "Legislative roles and missions of the Department of the Air Force." Unpublished thesis, U.S. Air War College, 1956.

Pringle, B.M. "The relationship between elements in the social situation and combat leadership behavior in the United States Air Force," 1953 (Ph.D., University of Southern California).

Richlin, M. "Behavioral analysis and evaluation of air force combat leadership." New York: New York University, 1955 (Ph.D. dissertation).

Roff, M. "A study of combat leadership in the Air Force by means of a rating scale: group differences." Journal of Psychology, 1950, 30, 229-239.

United States Air Force Academy, Department of Behavioral Sciences. "Incidents of leadership in combat." 1962.

Van Fleet, D.D. "Military leadership education: What war college journals seem to suggest." Proceedings of the Ninth Symposium on Psychology in the Department of Defense, 18-20 April 1984, United States Air Force Academy, Colorado Springs, Colorado, pp. 333-337.

Summary

The electronic bibliography on military leadership described in this paper is available on floppy disks for IBM PCs and compatibles. The bibliography is an up-dated version of that contained in Van Fleet and Yukl's (1986) monograph. In its present form, it has 4,101 entries. Although extensive, the bibliography can be quickly searched using the software provided (FYI 3000 Plus).

It is hoped that the availability of this electronic bibliography will lead to more integrative work on military leadership than has taken place in the past. The bibliography contains references to all branches of the service and related forms of organization (references containing the word army total 534; those with air and force 224; those with navy or naval 283; those with marine but not the merchant marine total 87; those with coast guard or coast artillery total 15; those related to the merchant marine total 5; and there are even 2 which deal with police organizations other than military or air police).

References

Bush, Vannevar (1945) "As We May Think", The Atlantic Monthly, 176, July, pp. 101-108.

Kimmel, J. R. FYI 300 Plus. Austin, TX: FYI, Inc., 1986.

Nelson, Theodore H. (1965) "A File Structure for the Complex, the Changing and the Indeterminate." Proceedings of the 20th ACM National Conference, pp. 84-100.

Van Fleet, D. D. & Peterson, T. O. (1986) "Toward An Electronic Bibliography of Military Leadership", Proceedings of the Tenth Symposium: Psychology in the Department of Defense, 16-18 April, pp. 633-637.

Van Fleet, D. D. & Yukl, G. A. (1986) Military Leadership: An Organizational Behavior Perspective, Greenwich, CT: JAI Press, 1986.

Using Set Correlations to Quantify Leader-Follower
Role Transactions in Two Army Reserve Units

Arthur B. Sweney, Ph.D.
Wichita State University
and
Colonel Donald Hackett, Ph.D.
United States Army Reserve
Wichita State University

Abstract

A sample of Army reservists was matched into 64 hierarchical
dyads to test the applicability of set correlations to measure
the degree of relationship between categories of variables
within and between dyadic partners. Role behaviors and other
perceptions were measured with the Role Reaction Battery.
These measures were correlated with satisfaction and efficiency
ratings as output-payoff measures and with education, gender,
tenure, and age as background qualification measures. The
intersection of sets were classified and found to be differen-
tially significant.

Since George Mead (1934) formally launched role psychology and Thibaut &
Kelly (1959), Blau (1964), and Jacobs (1972), identified the exchange process
transacted between leaders and followers, there have been numerous attempts
to measure the nature of the linkage in vertical dyads. Meaningful progress
has been demonstrated by Graen & Cashman (1975) with their "Role making
Model", by Sweney & Fiechtner (1976) with the Response to Power Role model,
and by Hersey and Blanchard (1982) using their Situational Leadership Theory.
Even the more recent and more empirically demonstrable Role Reaction Model,
(Sweney 1986), focused on too small a segment of the potential interactive
space between superordinates and subordinates. Out of frustration, such
investigators as Hunt and Osborne (1982), Hammer & Turk (1987) have had to
develop greater specificity in identifying the nature of the transaction
existing between the leader and the followers.

Many meaningful measures have been developed to study the theories
proposed, Stogdil's (1959) Leadership Behavior Description Questionnaire,
Fleischman's (1957) Leadership Opinion Questionnaire, Hersey and Blanchard's
Life Cycle Leadership Measures and Yukle and Nemeroff (1979) Leadership
Behavior Scales. But they all focus on the leaders and leave the followers
unmeasured. Since roles define behaviors chosen to be displayed in specific
positions the same behavioral trait displayed by a leader would represent a
different role when displayed by a follower. Both the RPM (Sweney 1970) and
the RRT (Sweney and Fiechtner 1976) take this into consideration in
identifying matching and complimentary superordinate and subordinate role
patterns and in defining intersecting relationships between these dyadic
partners.

The validation of the Role Reaction Model has been reported in both
Sweney and Sweney (1980) and V. A. Sweney (1986) with substantially different
populations. The current study is directed toward putting these role rela-
tionships into the larger context of the other possible transactions being
conducted.

Method

Subjects

The research was conducted upon 64 superordinate-subordinate dyads sampled from two Army reserve units during their monthly meetings. The units were chosen because their functions and organizational structures were radically different and, hence, their combined attributes would more nearly represent other reserve units.

Instruments

The Role Reaction Model illustrated in Figure 1, was measured in this study by a battery of four-forced choice ranking instruments. The Subordinate Motivation Rating provided fourteen five-response items which allowed the superordinate to rate each subordinate on five subordinate roles, the Critic, Rebel counteractive, Rebel interactive, Ingratiator counteractive and Ingratiator interactive. The superordinate also identified his/her own role behaviors when associated with each subordinate using the Superior Subordinate Reaction Test (Superordinate Form). When scored this identified his/her differential usage of the five leadership roles, Equalitarian, Authoritarian interactive, Authoritarian counteractive, Permissive counteractive, and Permissive interactive. The Cronbach alphas for these scales range between .74 and .96.
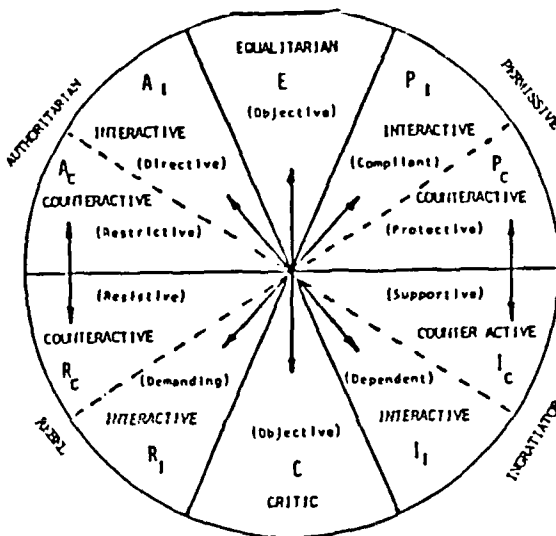
The subordinates responded on the Supervisors Purpose Rating to their perception of their superordinate's role in the five leadership areas, and they identified indirectly their own role pressures on the five subordinate roles using the Superior Subordinate Reaction Test (Subordinate Form). The Cronbach alpha reliabilities coefficients for these ten scales range between .80 and .91.

Demographic data were gathered from both dyadic partners. These included age, education, military efficiency ratings, tenure both on active and reserve status, ratings of their dyadic partner, and general satisfaction with reserve duty

Procedures

The tests were scored and then matched by dyads. In this way roles, other perceptions, qualifications and outcome measures could be correlated both within and across subjects. Since groups of variables constituted each of these data categories, Cohen and Cohen (1983) set correlations were calculated using the Setcoran program developed by Eber (1987). The classifications of the interactions of sets are defined in Table 1.

Table 1: The Classification of Dyadic Transactions
Based upon Intersections of Sets of Variables

Variable sets (# of var.)	Superordinate				Subordinate		
	Other Percept.	Own Roles	Background Qualificat.	Outputs Payoffs	Other Percept.	Own Roles	Background Qualifica.
Superordinate Roles (5)	Role Adaptat. by Super						
Superordinate Qualifications (6)	Background Anchoring by Super.	Background Role Press. on Super.					
Superordinate Outputs (3)	Climate Payoffs for Super	Role Payoffs for Super	Qualificat. Payoffs for Super				
Subordinate's Perceptions (5)	Mutual Percept.	Behavior Targeting by Subor.	Qualific. Targeting by Subor.	Standards Targeting by Subor.			
Subordinate's Roles (5)	Behavior Targeting by Super	Role Inter- action	Qualific. Role Press. on Subor.	Role Contrib. by Subor.	Role Adapt. by Subor.		
Subordinate's Qualifications (6)	Qualific. Targeting by Super	Qualific. Role Press. on Super.	Qualific. Compat- ability	Experience Contrib. by Subor.	Background Anchoring by Subor.	Background Role Press. on Subor.	
Subordinate Outputs (3)	Standards Targeting by Super	Leadership Contribut. by Super	Experience Contribution by Super	Output Inter- Dependence	Climate Payoffs for Sub.	Role Payoffs for Sub.	Qualificat. Payoffs for Sub.

Results

From the Pearson Product moment correlation matrix (available from
authors) it can be seen that the correlations within subjects and across
dyads for the Role Reaction Model scales were largely significant in the
directions predicted by the model and found in previous studies.   These and
other transactional processes, not predicted, are reported in Table 2 using
squared set correlations and significant levels based upon Rao's F test and
rounded up to .0001 if greater.
      It can be seen that Role Adaptation by both the superordinate and
subordinate were highly significant.   Superordinate and subordinate
Background Role Pressures barely missed significance. Neither the Climate
Payoffs nor Role Payoffs for either superordinates or subordinates was
significant.   However, Qualification Payoffs were significant in both
cases.

Table 2: Squared Set Correlations between Categories of Data

Variable Sets	Other Percept.	Own Roles	Background Qualificat.	Output Payoffs	Other Percept.	Own Role	Background Qualificat.
Superordinate Roles	$R^2 = .67$ $p = .0001$						
Superordinate Qualifications	singular	$R^2 = .44$ $p = .06$					
Superordinate Outputs	$R^2 = .27$ $p = .26$	$R^2 = .17$ $p = .75$	$R^2 = .67$ $p = .0001$				
Subordinate Perceptions	$R^2 = .56$ $p = .007$	$R^2 = .38$ $p = .32$	$R^2 = .68$ $p = .0002$	$R^2 = .80$ $p = .0001$			
Subordinate Roles	$R^2 = .65$ $p = .0001$	$R^2 = .37$ $p = .39$	$R^2 = .56$ $p = .02$	$R^2 = .53$ $p = .0001$	$R^2 = .72$ $p = .0001$		
Subordinate Qualifications	$R^2 = .55$ $p = .04$	$R^2 = .43$ $p = .36$	$R^2 = .56$ $p = .02$	$R^2 = .35$ $p = .12$	$R^2 = .58$ $p = .0002$	$R^2 = .52$ $p = .068$	
Subordinate Outputs	$R^2 = .29$ $p = .17$	$R^2 = .21$ $p = .56$	$R^2 = .30$ $p = .31$	$R^2 = .25$ $p = .05$	$R^2 = .25$ $p = .34$	$R^2 = .19$ $p = .63$	$R^2 = .41$ $p = .03$

Mutual Perceptions, Qualification Compatibility, and Output Interdependence were all significant but Role Interactions were not. The set correlation between the superordinate's perception and the subordinate's roles was significant indicating successful Behavioral Targeting of Perceptions by the superordinate. The reverse, however, was not true so the subordinates did not base their own perceptions upon leader's claimed behavior. The roles played by subordinates contributed significantly to the payoffs realized by their bosses but the roles played by the bosses did not contribute significantly to the payoffs realized by the subordinates. Perceptions by subordinates seem to be targeted on the superordinate's qualifications and outputs while superordinate's perceptions are targeted on the subordinate's role behaviors and other perceptions.

Discussion

The relationships found in this study were neither as strong nor as orderly as those found in the other organizations studied where the intensity of contact was stronger. Some time had to be taken with the reservists for them to reaffirm to whom they reported and, hence, who they should rate. This attests to the loose linkages involved.

This was a good design to test the interpretability of set correlations for viewing complex interpersonal interactions. The size of the sample was too small, however, to provide generalizations which would not require future testing.

References

Blau, P. M. (1964) *Exchange and Power in Social Life*. New York: John Wiley and Sons, Inc.

Cohen, J. & Cohen, P. (1983) *Applied Multiple Regression Correlations Analysis for the Behavior Sciences*, 2nd Edition. Hillsdale, N.J.: Lawrence Erlbaum Associates, Publishers.

Eber, H. (1985), *Setcoran*. Atlanta, Ga: Psychological Resources

Graen, G. & Cashman, J. (1975) A role-making model of leadership in formal organizations. In J.G. Hunt & L.L. Larsen (Eds) *Leadership Frontiers*, Kent Ohio: Kent State University Press.

Hammer, T. H. & Turk, J.M. (1987) Organizational determinants of leader behavior and authority, *Journal of Applied Psychology*, 72, (4), 674-682

Hersey, P. & Blanchard, K.H. (1969) Life cycle theory of leadership. *Training and Development Journal*, 23, (5), 26-34

Hunt, J.G. & Osburn, R.N. (1980) A multiple influence approach to leadership for managers. In J.S. Stinson & P. Hersey (Eds) *Perspectives in Leader Effectiveness*. Athens, Oh: Center for Leadership Studies, Ohio University

Jacobs, T.O. (1970) *Leadership and Exchange in Formal Organizations*. Alexandria, Va: Human Resources Research Organization.

Mead, G. H. (1934) *Mind, Self, and Society*. Chicago: University of Chicago Press.

Sweney, A. B., (1970) Organizational power roles. *Professional Management Bulletin*, 10, 5-13

Sweney, A.B. & Fiechtner, L.A. (1976) Factors in subordinate-superordinate interaction, *Multivariate Experimental Clinical Research*, 2, 93-100

Sweney, V. A. & Sweney, A. B. (1980) Validating the role reaction model for predicting superior-subordinate relationships. *Proceeding: Psychology in the Department of Defense Seventh Annual Symposium*. Colorado Springs.

Sweney, V. A., (1986) Role perceptions and role adaptations in hierarchical dyads, *Proceedings: Psychology in the Department of Defense, Tenth Annual Symposium*. Colorado Springs.

Thibaut, J.W. & Kelly, H. H. (1959), *The Social Psychology of Groups*. New York: John Wiley and Sons.

Yukle G. A. & Nemeroff, W. F. (1979) Identification and measurement of specific categories of leadership behaviors: A progress report. In J.G. Hunt & L.L. Larson (Eds) *Cross Currents in Leadership*. Carbondale, Il: Southern Illinois University Press.

ARMY CULTURE, COMMAND CLIMATE, AND COMBAT POTENTIAL

LTC Faris R. Kirkland, Ph.D., USA-ret.
Department of Military Psychiatry, Walter Reed Army Institute of Research

Abstract

Longitudinal survey and observational research on human relations
was conducted in 137 Army companies. Supportive command climate proved
to foster the ability of subordinate leaders to command effectively.
Elements of Army culture--competitiveness, fear of responsibility, and
condemnation of fraternization--interfered with battalion commanders'
efforts to create supportive command climates. Modification of these
cultural components is necessary to enable units to develop the cohesion
and competence necessary to fight successfully in AirLand and low/mid-
intensity warfare.

Descriptions of future wars suggest that small units will have to operate
autonomously for prolonged periods under extreme stress. Characteristics of units
that can sustain themselves psychologically and can continue to fight in the face
of isolation and continuous violence have begun to emerge. Since the 1940s a
growing body of research has linked cohesion in small units with resistance to
combat stress breakdown and to effective performance in combat (Stouffer, et al.,
1949; Shils and Janowitz, 1948; Marshall, 1947). Recent research has documented
the characteristics of high performance units (Malone, 1986; Simonsen, et al.,
1985). Gal (1986), Henderson (1985), Malone (1983), and Van Creveld (1982) have
described the leadership likely to produce cohesion and high performance.

The department of Military Psychiatry at the Walter Reed Army Institute of
Research has, over the past eight years, been investigating relationships between
policy, command, and leadership on the one hand, and resistance to stress break-
down, cohesion, and performance on the other (Kirkland, 1987, 1988; Marlowe, 1985,
1986a, 1986b, 1987; Ingraham, 1984; Manning & Ingraham, 1984). In the process the
research team identified an additional variable--Army culture--that distorts
expected causal relationships. Army culture is the set of ways in which soldiers
handle situations and relate to each other. The culture is carried by and influ-
ences the activities primarily of career soldiers. It therefore operates in the
domains of command, staff, and leadership--between policy and the private soldiers.
This paper is a report on the effects of Army culture on leadership, command
climate, and unit effectiveness.

Method

The research design includes surveys, interviews, and participant observation
in a large number of combat units. Surveys were administered to members of 137
company-sized units five times during a three-year period. The surveys addressed
social supports in units, confidence in leaders and equipment, and attitude toward
going into combat with the unit (Marlowe, 1985, 1986a, 1986b). Structured inter-
views reached representatives of all echelons in more than 60 battalions. Approxi-
mately 25 percent of the soldiers in each battalion participated in a group or
individual interview. The interviews focussed on leader behavior, leader-follower
relationships, and expectations of success in combat. Participant observers lived
with units in garrison and during simulated combat operations. They recorded

leader behavior, interpersonal dynamics, responses to crises, and performance of mission-related tasks. Quantitative data from the surveys identified issues for detailed investigation by the interviewers and observers, and the qualitative data informed interpretation of the survey results.

Results

The research team found that the first-term soldiers readily accepted discipline, took an interest in learning to be effective combat soldiers, and identified with the unit mission. By correlating observed leader behavior at squad, platoon, and company levels with observed performance and with survey measures of cohesion, it was possible to identify specific acts and attitudes on the part of leaders that brought these willing soldiers together in committed, competent units.

The effective leaders were those who knew a lot about their profession, wanted to learn more, and were eager to experiment with their own and their subordinates' ideas. They accepted the risks and responsibilities that went with empowering junior leaders--entrusting them with missions and granting them discretion. They trusted their subordinates, respected them as junior colleagues, shared their professional interests with them, and were active in assuring their professional, personal, and familial welfare. These kinds of leadership behavior are currently referred to as positive leadership. Those few leaders who practised positive leadership had exceptionally cohesive and competent units. The extent to which leaders lacked technical competence, centralized control, distrusted their subordinates, and held themselves remote from and indifferent to their subordinates was closely correlated with fragmentation, indiscipline, and poor performance on mission-related activities. Most leaders expressed approval of positive leadership; that most did not employ it when it was manifestly in their interest to do so provoked a fresh analysis of the data.

The results of the reanalysis showed three things. First, a leader was much more likely to use positive leadership if his boss was a positive leader. Second, positive leadership is exceptionally stressful because it demands a higher degree of professional knowledge, candor, and personal accessibility than does distant or authoritarian leadership. Also, trusting and empowering subordinates entails risk and uncertainty. Third, one of the principal purposes of Army culture is to protect professional soldiers from exposure to the stresses of exercising positive leadership.

Discussion

Together, these three findings suggest that a leader can only afford to take on the stresses associated with being a positive leader if he enjoys respect, trust and concern from his superior. Conversely, a leader whose superior is suspicious, authoritarian, and micro-managing has to devote his energies to protecting himself; he has none left to be a positive leader for his subordinates. Battalion commanders who viewed their subordinates as adversaries whom they had to watch and drive had indecisive, discouraged company commanders. Battalion commanders who saw their role, and that of their staffs, as supporting their company commanders had cohesive and competent companies. Support took three forms: purposing, respect and staff responsiveness.

Purposing is the process by which the commander articulates a realistic set of objectives for his battalion, and cleaves to it. In terms of the colonel's behavior, it means that he makes the hard choices about which requirements to emphasize, and which ones to ignore. By setting the priorities himself, he takes

the heat for non-performance of the peripheral requirements. He also shows the members of his battalion that he believes in the objectives he has laid out, and has the courage to stand up for them.

Respect refers to the battalion commander treating his subordinate commanders as worthy junior colleagues in whom he reposes trust and confidence, and for whom he will stand up when they err in the process of learning to command. The most successful battalion commanders were like older brothers with their captains. The colonels allowed their captains substantial autonomy in running their companies, yet were ready to discuss problems or answer questions. In their after-action critiques, the colonels invited their captains to discuss ways in which he and his staff could have contributed more effectively to the operation as well as what the captains could have done better.

The third form of support the effective battalion commanders provided was a competent staff dedicated to assisting the subordinate units. The most important contribution the staff could make was to integrate intelligence, logistical, and operational data on FTXs to assure that companies had appropriate missions, were fully informed, and had adequate logistical and fire support. The second most important function of the staff was responsive attention to the companies' needs for personnel, training, and logistical support. The performance of his staff was a statement of the colonel's commitment to the success of his companies.

Few of the battalion commanders in the units studied were able to provide any of the support necessary for their captains to command effectively, and most provided none. Here Army culture operated to vitiate both policy and the dedication of battalion and company commanders. The research team identified several facets of Army culture that, by degrading the effectiveness of command, degraded the war-fighting capacity of the units. These were emphasis on competition, fear of responsibility, and condemnation of fraternization.

The research team found that the competitive tradition, places a premium on looking good rather than being good. It is the foundation of the "Can Do" philosophy, which substitutes fantasy for reality in assessing the capabilities of units and the assignment of tasks. Competitiveness drives commanders to waste their troops' energies on demonstrations, turn their staffs into VIP briefing teams, and micro-manage their units to create an impression that they have capabilities they have not had time to develop substantively. The competitive tradition does not make sense now when battlefield teamwork has become the product of independent decisions by leaders of small units.

Fear of responsibility is most frequently manifest in the reluctance of com-manders to set priorities. Most battalion commanders passed requirements down to their companies with a number one priority. The captains had to decide which requirements to emphasize and which ones to ignore. This procedure has the double advantage of sparing the senior commander the burden of making decisions, and of allowing him to dodge the bullet if some of his captains' choices do not meet with the approval of the colonel's superiors.

The tradition of condemning intimacy across ranks--fraternization--has been adaptive for generations of leaders who feared they were not competent to hold their ranks. It gives them distance and rituals behind which they can hide their shortcomings from their subordinates. The research team observed that cohesive, competent units showed consistently that leaders and followers were close, forth-right, and friendly with one another. In fragmented, floundering units, leaders were remote, punitive, and disparaging toward their subordinates.

The maladaptive aspects of Army culture arise from insecurity among military professionals--the beneficiaries and the propagators of the culture. Generations of career soldiers have fashioned the Army culture to protect them from having to deal with the technical complexities of their branches, from accepting responsibility, and from becoming close with their subordinates. But in AirLand and low/mid-intensity scenarios, units are most likely to remain cohesive and combat effective when they can trust their commanders to do everything in their power to care for and support them, and never abandon them on the battlefield. The Army can only expect to win if it fosters the development of professionals worthy of that trust.

It takes strenuous effort, and often many years, to change cultural tenets that protect the members of a culture from psychological stress. Some soldiers will not be able to function without the old Army culture; their need for it is too strong. As one Division Command Sergeant Major said: "When I find an NCO who doesn't have the guts to trust his people, I tell him to start looking for a new home." But if assumption of responsibility and concern for subordinates' sense of security start at the top, it may be possible to restructure human relations in the Army in a relatively short time. It will take a serious effort by the senior management of the Army, followed by an equally serious effort on the part of intermediate commanders, to help their subordinates feel secure by trusting, respecting, and supporting them.

References

Gal, R. (1986). A portrait of the Israeli soldier. Westport, CT: Greenwood Press.

Henderson, W.D. (1985). Cohesion: The human element in combat. Washington, DC: National Defense University Press.

Ingraham, L.H. (1984). The boys in the barracks. Philadelphia: Institute for the Study of Human Issues.

Kirkland, F.R. (1987). Leadership in COHORT companies. Washington, DC: Walter Reed Army Institute of Research.

Kirkland, F.R. (1988). Promise and problems in commanding COHORT battalions. Washington, DC: Walter Reed Army Institute of Research.

Malone, D.M. (1986). An army of excellence. Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Malone, D.M. (1983). Small unit leadership: A commonsense approach. Novato, CA: Presidio.

Manning, F.J., & Ingraham, L.H. (1984). An investigation into the value of unit cohesion during peacetime. (WRAIR-NP-83-5). Washington, DC: Department of Military Psychiatry, Walter Reed Army Institute of Research.

Marlowe, D.H. (Ed.). (1985, November). New manning system field evaluation: Technical report no. 1. Washington, DC: Walter Reed Army Institute of Research.

Marlowe, D.H. (Ed.). (1986a, March). New manning system field evaluation: Technical report no. 2. Washington, DC: Walter Reed Army Institute of Research.

Marlowe, D.H. (Ed.) (1986b, June). New manning system field evaluation: Technical report no. 3. Washington, DC: Walter Reed Army Institute of Research.

Marlowe, D.H. (Ed.). (1987, September). New manning system field evaluation: Technical report no. 5. Washington, DC: Walter Reed Army Institute of Research.

Marshall, S.L.A. (1947). Men against fire. Gloucester, MA: Peter Smith. (Reprinted 1978).

Shils, E.A., & Janowitz, M. (1948). Cohesion and disintegration in the Wehrmacht in World War II. Public Opinion Quarterly, 12 , 280-315.

Simonsen, J.A., Frandsen, H.L., & Hoopengardner, D.A. (1985). Excellence in the combat arms. Military Review, 65 (6), 19-29.

Stouffer, S.A., Devinney, L.C., Star, S.A., & Williams, R.M. (1949). The American soldier. Princeton: Princeton University Press.

Van Creveld, M. (1982). Fighting power. Westport, CT: Greenwood Press.

Principal Components Analysis For
Multicollinerarity and Homogeneity
of Transactional and Transformational
Leadership Dimensions

Ruediger Mueller
William E. Rosenbach
Gettysburg College

Abstract

Principal components analysis was performed
to test the abbreviated version of the Bass Multi-
factor Leadership Questionnaire.   Data were
collected from 1304 subordinates of 171 business,
church, and fire service leaders.   Results are
discussed in terms of the appropriateness of the
abbrevated questionnaire for use in leadership
development programs.

In his Pultizer Prize winning book, <u>Leadership</u> (1978), James
McGregor Burns identified and described two basic types of
political leadership, the transactional and the transformational.
He described the relations of most leaders and followers as
transactional - leaders approach followers with an eye to
exchanging one thing for another: jobs for votes, or subsidies
for campaign   contributions.   Burns described transforming
leadership as more complex, but also more potent.   The trans-
forming leader recognizes and exploits an existing need or demand
of a potential follower. But, beyond that, according to Burns,
the transforming leader looks for potential motives in followers,
seeks to satisfy higher needs, and engages the full person of the
follower. The resulting transforming leadership is a relation-
ship of mutual stimulation and elevation that converts followers
into leaders.

Bass made the most significant contribution to the study of
leadership in the past two decades by extending Burns' concepts
to apply to organizational management.   Bass (1985) described
transactional leaders in their relations with subordinates as:
(1) recognizes what it is they want to get from their work and
tries to see that they get what they want if their performance
warrants it (2) exchanges rewards and promises of rewards for
their efforts, and (3) is responsive to their immediate self-
interests if they can be met by their getting the work done.
Transactional leadership is based on the expectancy theory of
motivation (Vroom, 1964) and the path-goal theory of leadership
(House, 1971).   The leader recognizes the role the follower must
play to attain outcomes desired by the leader.   The leader
clarifies the role and provides the follower with the confidence
necessary to carry it out and meet the objectives. In parallel,
the leader recognizes what the follower needs and clarifies for
the follower how those needs will be fulfilled in exchange for

the followers' satisfactory effort and performance. (Bass, 1985, p. 13). Properly implemented, transactional leadership is effective and desirable in the appropriate situation.

Transformational leadership involves strong personal identification with the leader, joining in a shared vision of the future, and going beyond the self-interested exchange of rewards for compliance. The transformational leader motivates followers to perform beyond expectations by: (1) raising an awareness of the importance and value of designated outcomes, (2) influencing followers to transcend their own self-interests, or (3) altering or expanding followers' needs on Maslow's hierarchy of needs (Bass, 1985).

Hatter and Bass (1987) warn that the constrasting of transactional leadership with transormational leadership should not be construed to mean the models are unrelated. Transformational leadership may be described as a special case of transactional leadership since both styles are linked to achieving a goal or objective. The two models differ in the manner by which the leader motivates subordinates and the types of goals that are set. The transformational leader may demonstrate transactional leadership at times, but also use symbolism or imaging to elevate the importance of increased effort for accomplishing the mission, which serves as a motivator itself. In the course of working to accomplish the mission, followers may enhance their own development (Hatter and Bass, 1987, p. 5).

Transformational leadership differs from transactional in that transformational leaders do recognize follower's current needs and objectives but they attempt to elevate those needs. Motivation beyond expectations is achieved by raising the expecttations followers have for their own individual needs and performances. Transformational leaders also differ from transactional in that the leader attempts to develop followers into leaders. Transformational leaders improve the capabilities of followers to solve their own problems as well as organizational problems (Avolio and Bass, 1987).

Bass (1985) developed the Multifactor Leadership Questionnaire (MLQ) for measuring those dimensions which comprise transactional and transformational leadership. The construct validity of the 73 item MLQ was accomplished by Bass (1985) with data from 104 military officers who completed the instrument describing their immediate supervisors. Six additional items on the MLQ dealt with the leaders' effectiveness and their followers satisfaction with their style. Clover (1986) reduced the 79 item MLQ to an instrument with 36 items to measure transactional and transformational dimensions and seven items to measure leader effectiveness and follower satisfaction with the leader. When performing management development projects in organizations, the use of lengthy questionnaires often hampers candid and commited participation of the members. The purpose of this study was to test the validity of an abbreviated version of the MLQ for use for leadership development programs.

Method

The data was collected from three groups, fire service executives (firechiefs), industry leaders (CEOs, presidents, and

vice-presidents), and protestant pastors (clergy). Each subject was rated by seven to ten subordinates in a total sample of 171 subjects and 1304 ratings. The sample of firechiefs and industry leaders was collected over a period of six months in a series of leadership development projects by the authors. The sample of clergy was collected for a dissertation project following the guidelines of the authors. The data were collected using the Leadership Description Questionnaire, a somewhat revised version of Clover's (1986) abbreviated version of the Multifactor Leadership Questionnaire (Bass, 1985).

Respondents described the frequency of transactional and transformational leadership behaviors of their leaders on a scale of one to five which ranged from one, not at all, to five, frequently if not often. Transactional leadership was determined by three components, contingent reward, management by exception, and laissez faire. Transformational leadership was defined as a combination of charisma, individualized consideration, intellectual stimulation, inspriation, and motivation beyond expectation. Leader effectiveness was measured on a scale of one, not effective, to five, extremely effective. The effectiveness measurement included overall effectiveness of the unit, effectiveness of leader in representing unit to higher authority, effectiveness in meeting job related needs of subordinates, effectiveness in meeting organizational requirements, and effectiveness in meeting personal needs of subordinates. Satisfaction with the leader as a person and satisfaction with the leader's style were similarly measured on a scale of one, very dissatisfied, to five, very satisfied.

The purpose of the analysis was twofold. It was intended to check for the presence of multicollinearity among the dimensions of transactional and transformational leadership as well as the outcome variables, leader effectiveness and follower satisfaction. The second objective was to check the individual dimensions of leadership themselves for consistency, i.e. would the items which account for each dimension, group together when a principal component analysis was performed.

The analysis for multicollinearity was performed to confirm the validity of regression analysis on the data performed elsewhere (Rosenbach & Mueller, 1988). Two measures both calculated in the context of principal components analysis, where used to measure the presence of multicollinearity. These were Variance Inflation Factors (VFI) and Condition Indices, (Belsley, D. A., Kuh, E. & Welsch, R. E., 1980). The Variance Inflation Factor is calculated as: $VFI_i = \frac{1}{(1-R^2_i)}$
where VFI_i = Variance Inflation Factor for Variable i,
and R^2_i = squared multiple correlation of variable i with the other variables.

Multicollinearity is presumed to be present, if any VFI exceeds 10 or if condition indices exceed 30 with the variance proportion of the corresponding item for two or more principal components exceeding .5. Analysis of the data reveals, that multicollinearity is not a factor when using the leader description questionnaire. For the transformational leadership dimensions all condition indices are less than 5, for four items

248

the condition index is slightly above 30, but in none of these cases is the variance proportion higher than .5 for two or more items. The findings for transactional leadership deminsions are even stronger, in this instance all VIF;s are less than 3 and no condition index exceeds or approaches 30. The same holds true for the outcome variables, effectiveness and satisfaction. Here VIF's and condition indices also remain well below the critical limits.

These results lead to two conclusions; first, any further analyses using regression or similar methods can be performed without fear of distorted results due to the presence of multicollinearity. The second conclusion may be even more important; a high multicollinearity implies that at least one item does not contribute substantially to explained variance above and beyond the variance explained by the other items, thus implying that it may be dropped from the analysis without substantial loss of explanatory power. However,this is not the case here, all items on the questionnaire contribute to explained variance and therefore should be retained.

The second part of the analysis intends to test the homogeneity of leadership dimensions and the outcome variables. Each dimension is measured by a set of 3 to 5 items in the questionnaire. The research question of interest is, do these questions really measure the same construct? For example, do all items referring to contingent reward really refer to the same underlying construct, namely contingent reward? The method used to answer the research question is, again, principal component anaylsis. While this method generates as many components (linear combinations of items) as there are items, it is expected, that a few components explain most variance, with the remaining components contributing little (usually in the low single digit range) to explained variance. These can usually be ignored for purposes of analysis. In the present case, one would expect to retain as many components as there are leadership dimensions and to find variables measuring the same dimension loading highest on the same component.

Results

The results of the analysis are somewhat inconsistent with our expectations. In the case of transformational leadership dimensions (five), the first five components account for only about 75 per cent of total variance. While this might still be regarded as acceptable, loadings defy expectations. Motivation beyond expectation (MBE) is measured by three items, all of which load strongly on different components; only one of these loads strongly on one of the first five components. As we would have expected, individualized consideration (four items) shows a similar pattern. In the case of charisma (four items), two items load strongly on the same component, but one which contributes only 1.1 per cent to explained variance. None of the four chairsma items loads on any of the first five components. Intellectual stimulation and inspiration (four items) each show a similar pattern.

In the case of transactional leadership which consists of three dimensions, the first three components account for only

249

about 50 per cent of variance. To account for 75 per cent of variance, seven components must be considered. Management by exception (four items) is the only dimension which loads strongest on one of the first three components. But no two of the items load strongest on the same component. Two of the items measuring laissez faire (five items) load strongest on the same component, which accounts, however, for only four per cent of overall variance. Contingent reward (four items) shows a similar pattern.

In case of the outcome variables the results look somewhat closer to expectations. Both satisfaction variables load strongest on the same component, albeit once again, one which contributes little (2.5 per cent) to explained variance. Only two of the five effectiveness variables load strongest on the same component.

Discussion

Conclusions from these findings must be drawn with care and should be regarded as only a first step in testing the abbreviated questionnaire. It appears that the dimensions of transactional and transformational are less homogeneous than one might expect, as indicated by the lack of consistent loadings on principal components. Bass' (1985) own findings with the MLQ are similar. In our opinion, the continued use of the abbreviated questionnaire is appropriate for management development efforts since our findings have shown that the items used are not redundant and satisfactorily measure the transactional and transformational leadership construct. As additional data are gathered in a variety of organizational settings, a more thorough analysis is suggested.

REFERENCES

Avolio, B.J. & Bass, B.M. (1987). Charisma and beyond. In J.G. Hunt (Ed.), Emerging Leadership Vistas. Boston: Lexington.

Bass, B.M. 1985. Leadership and performance beyond expectations. New York: The Free Press

Blanchard, K.H. & Johnson S. 1982. The one minute manager, New York: William Mossow

Burns, J.J. 1978. Leadership. New York: Harper and Row

Clover, W.H. (1986). Item reduction of the Multifactor leadership questionnaire. U.S. Air Force Academy Working Paper.

Hatter, J.J. & Bass, B.M. 1987. Supervisors' Evaluations and Subordinates' perceptions of transformational and trans-actional leadership. SUNY at Binghamton, School of Management Working Paper Series.

House, R.J. 1971. A path-goal theory of leadership effectiveness. Administrative Science Quarterly, 16, 321-338.

Rosenbach, W.E. & Mueller, R. (1988). Transactional and transformational leadership of business, church, and fire service executives. Under review.

Vroom, V.H. 1964. Work and Motivation. New York: John Wiley & Sons

Computer Games and Human Performance
David B. Porter
USAF Academy
Oxford University

This paper outlines the issues and concerns which led to the development of an arcade-type computer game for exploring human performance. Synopses of three early experiments are presented and discussed. Results from these experiments suggest dificiencies in conventional assumptions concerning verbal reports, task difficulty, side-task interference and the development of automaticity.

It is ironic that the type of human performance of greatest interest (i.e., skilled performers confronting crisis situations) is almost never studied in the laboratory. Fear of the additional variance introduced by differences in individual skills has led researchers to adopt experimental tasks so deficient in intrinsic motivational quality it can be safely assumed that none of the subjects would have relevant prior experience or will be motivated to "master" the task at hand. The stark simplicity of most laboratory tasks also reflects experimenters extreme concern with "control." Most researchers select "algorithymic" tasks which have straight forward solution paths. They then adjust performance levels by manipulating speed or accuracy requirements. "Heuristic" tasks, those which require strategies and involve higher-order processes (McGraw and McCullers, 1979) are generally seen as unsuitable for traditional experiments (Eysenck, 1983). Other critically important factors such as motivation and emotion have not only been ignored but specifically targeted for "control through minimalization." Many of our assumptions about human information processing are based on traditional laboratory experiments. Concern for the limitations this involves provided the primary impetus for this research.

Computer games offer experimental psychologists the unique opportunity to study the motivated performance of substantive (i.e., heuristic) tasks. Games are intrinsicly motivating (if they were not, noone would choose to play them). Games also teach skills and abilities which are applicable to many real world tasks. The computer can be a powerful colleague in addressing the inherent problem of inadequate measurement; it is capable of unobtrusively recording literally hundreds of measures every second. Such high-density measurement enables us to construct objective post-hoc models of the knowledge structures and information processes implied by regularities in subjects' performance. These models, in turn, can be compared with subject's verbal reports. Such models can also be used to identify the level and type of influences exerted by a variety of side tasks or environmental conditions.

Before reporting the experiments, it is necessary to explain the game. The Whale Game, could be learned by most subjects in 10 minutes but of 2,188 recorded experimental trials, only three were performed optimally.

The monitor displayed a cyan ocean containing several clusters of white icebergs (Figure 1). Subjects controlled the direction of movement of a blue whale by using the keyboard controls shown. The whale turned in the direction corresponding to the key pressed by the subject. It moved one space left, right, up or down about every 600 msec. It could not move diagonally or cross the screen boundary and if it swam through one of the icebergs, the iceberg disappeared. Players could score points by accomplishing either of two tasks: eating plankton or wrecking kayaks. The number of points for each task was designated before each trial. Before discussing the three priority conditions, it is important to point out substantive differences in the two game tasks.

The plankton task was simple but uncertain (very similar to a traditional "algorithymic" laboratory tracking task.) A small mass of green plankton began each trial in the center of the screen. It proceeded in a "random walk"

251
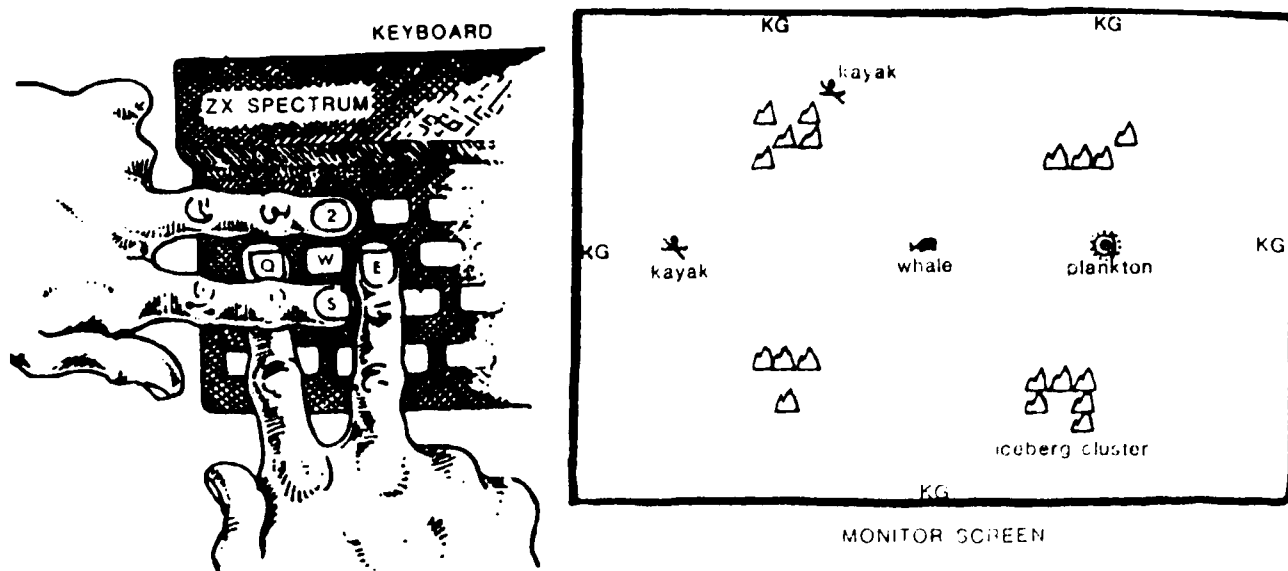
Figure 1. THE WHALE GAME - Controls and Display

to the right, drifting first to the top and then the bottom of the screen. It disappeared from the right screen border and reappeared at the extreme left border. It was relatively obvious how to hit the plankton, but doing this required skill and concentrated effort.

In contrast, the kayak task was complex but predictable (in McGraw and McCullers' (1978) terms a "heuristic" task). Red kayaks were "generated" from one of five locations along the screen border (labeled "KG" in Figure 1) at specified times during each trial. Once on the screen, kayaks followed one rule: they moved one space (horizontally, vertically or diagonally) toward the whale. Kayaks remained on the screen until one of two things happened: they encountered an iceberg and crashed or reached the whale and harpooned it. Subjects gained points for the former and lost an equal number of points for the latter. The number of kayaks present at any time varied from zero to three, but twenty were presented during each two-and-one-half minute trial.

A trial consisted of 218 computer "cycles" during which each character could move one space on the screen's covert 22 x 32 space matrix. Subject intention was manipulated directly by combining the two tasks under three different priority conditions. During the plankton priority trials, plankton was worth 100 points a bite and crashing or being harpooned by kayaks was worth plus or minus 10 points. Points were reversed for kayak priority trials. During equal priority trials, all events were worth 50 points. Priority was designated by instructions displayed on the screen before each trial. Both tasks required subjects to make discrete manual responses to visuo-spatial information displayed by the video screen. Differential effects on performance between priority conditions, therefore, could be attributed to intermediate (i.e., cognitive) rather than peripheral differences in the performance of the two tasks.

Several other differences in the two tasks are also important. Human performance of the plankton task can be approximated by three short lines of BASIC but the same degree of functional equivalence requires over 30 lines of BASIC for the kayak task. Subjects' verbal reports of the strategies they employed in performance of the plankton task were very consistent with their actual performance. However, for the kayak task, subjects' verbal knowledge of objectively instrumental behaviors was clearly deficient. Nearly all subjects rated the kayak task as being much more "difficult" than the plankton task (Porter, 1987). The synopses of the three experiments which follow suggest the utility of this approach to studying human performance.

252

control
5 letter memory load

fast   (605 msec)
medium (650 msec)
slow   (695 msec)

control
randomization (trials 1-9)
randomization (trials 10-18)
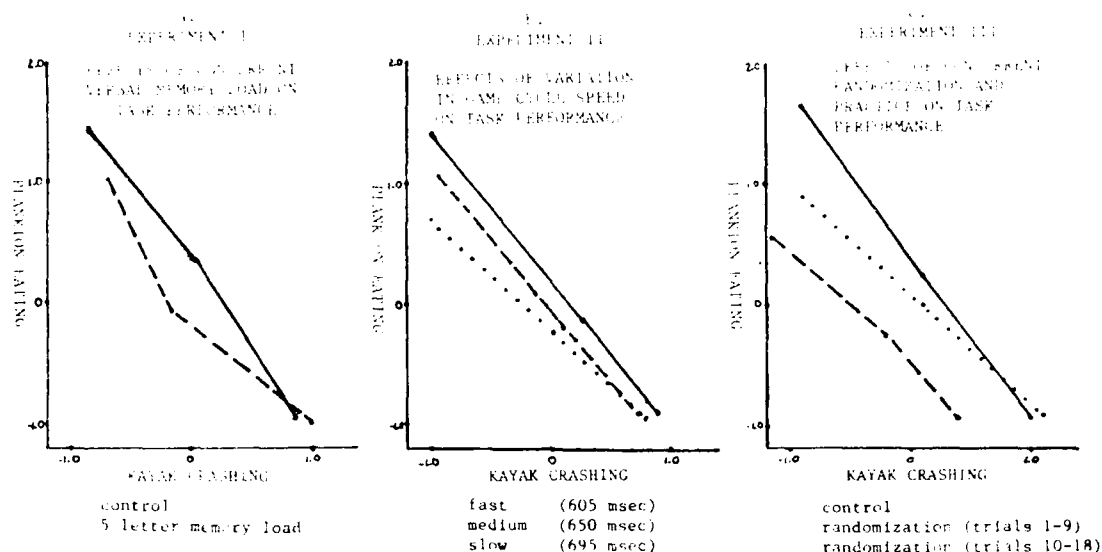
Figure 2. Differential Effects on Performance of Game Subtasks

Experiment 1

Ten male and ten female subjects from the Oxford Subject Panel were randomly selected. After performing a four choice reaction time task and whale control training, subjects were allowed nine practice trials before the experiment began. Priority instructions (i.e., plankton, equal or kayak) changed each successive trial. After a short break, the verbal side task was introduced to the subjects. Subjects were required to use subvocal rehearsal to retain 5 letter strings for periods of 25 seconds. At the beginning of the trial, the experimenter would say the 5 letters chosen from the set: F,H,J,L,M,Q,R,S,Z. Twenty-five seconds later, the experimenter would tap the desk top and the subject would respond with the appropriate letter or letters. The experimenter would then give the subject another set of letters to rehearse for 25 seconds. This continued until the end of the trial. Theoretically, this verbal side task should have occupied subject's articulatory loop and thus decreased their capacity for processing other verbal material (i.e., self-talk or self-instruction.) Each subject performed the priority with each memory load twice. Orders of presentation were counterbalanced. After each trial, subjects recorded their score on each game task and were told their score on the verbal side task by the experimenter.

There were no significant differences in subjects' performance of the memorial side task across priority conditions. Figure 2a contains average performance operating characteristics (Wickens, 1984) for the control (solid line) and the 5-letter memory load conditions (broken line). The scales for both axes are standard scores derived for each subject from their performance across all trials. As is apparent from the POCs, the memory load caused significant decrements in subjects' performance during both the plankton and equal priority trials. The effect of the 5-letter memory load on the performance of the kayak task, however, actually showed a slight increase. This was curious because subjects nearly unanimously identified this task as being the most difficult.

Experiment 2

Experiment 2 examined two separate influences: concurrent peripheral verbal activity and processing time available. Experiment 1's results could be interpreted as reflecting the greater susceptibility of the plankton task to interference from motor activity. This alternative was to be investigated by using a side task which required continuous verbal motor activity but

254

minimal intermediate processing: continuous fixed-order articulation. This experiment also sought to determine the effect of additional processing time by surreptitiously varying the game's cycle speed. This manipulation was designed to test the hypothesis that the differences obtained in the previous experiment were merely artifacts of subject's imperfect knowledge of the cognitive demand of the two tasks. It may have been that subjects garnered extra resources for accomplishing the kayak task in the dual task situation because they knew it was more difficult. This may not have occurred for the plankton task because subjects assumed it was very easy. The time variation selected (45 msec cycle) was not noticeable, thus different effects on the two game tasks could not be caused by conscious strategies.

Twenty-four subjects from the Oxford Subject panel who had not participated in the previous experiment were randomly selected and assigned to counter balancedorders of presentation of game speeds (with average cycle times of 605 msec, 650 msec and 690 msec) and priorities (i.e., plankton, equal, or kayak). For the peripheral verbal side task, subjects were to continuously repeat "left, right, up, down" at least once every 4 seconds at a volume sufficient for the experimenter to hear and record. Half the subjects verbalized on odd numbered trials and the other half verbalized on even trials. Each subject completed 18 trials.

There were no significant differences in the number of verbalizations by priority condition. Although the verbal side task did have a slight effect of the number of keyboard inputs subjects made, it did not significantly affect subjects' performance of either the plankton or kayak task. Time on the other hand had several significant effects. The POCs shown in Figure 2b reflect subjects' average performance at the three game speeds. Additional processing time enabled significant increases in performance of the plankton task. During equal priority trials, performance also benefitted from additional processing time. Additional processing time apparently also led to a shift in subjects' bias toward the kayak task. This is ironic because the additional processing time clearly did not improve subjects' performance of the kayak task during kayak priority trials.

Experiment 3

Results reported thus far could be explained very simply if the lack of negative effects on the kayak task were caused by an inherent lack of sensitivity in the task itself. (The fact that the performance of this task varied consistently with changes in practice and priority instruction argues strongly against this explanation.) Experiment 2 showed that peripheral motor activity did not affect either task and Experiment 1 suggested that a side task which involved the articulatory loop would affect performance of the plankton task but not the kayak task. What was needed was a demonstration of a side task which would cause significant decrements in both tasks. The generation of random sequences, hypothetically involves the operation of the central executive portion of Hitch and Baddely's (1976) working memory. It was also of interest, to see if the constant relation-structures underlying the kayak task would permit this task to become automatized more readily than the simple and "easy" but uncertain plankton task.

Twelve male and twelve female subjects who had not participated in previous experiments were randomly selected from the Oxford Subject Panel. As in Experiment 1, subjects performed four-choice reaction time tasks, whale control training and six practice trials before the experimental trials began. Two separate verbal side tasks were used. Both side tasks required subjects to respond with a cardinal direction (or number) with each click of a metronome placed atop the video screen (i.e., once every 1.5 seconds). In one condition, subjects made their responses in a designated, fixed order. It was assumed that this would occupy only subject's articulatory loop. In the other experimental condition, subjects were to respond randomly. Random number generation was discussed with each subject and then each produced two strings

of 100 numbers which were analyzed for randomness before the start of the experimental trials. Again each of the subjects completed all three priorities under each of the three conditions (control, fixed-order and random-order) twice. Orders of completion were counterbalanced.

The results from the fixed-order concurrent articulation task were nearly identical to those obtained in Experiment 1 with the memorial side task. (Significant decrements in plankton and equal priority trials, but no effect on the "more difficult" kayak task). The effects of the other side task, random generation are shown in Figure 2c. The control POC shows the average performance of all 24 subjects across the 6 control trials. Two separate POCs show the effects of concurrent randomization during the first 9 trials and again during the second 9 trials. The decrements in performance during the first 9 trials appear to be nearly equal across all three priority conditions. However, the POC shows a pronounced lateral shift to the right during the second set of nine trials. This is graphic evidence of the fact that despite subjects' rating of the kayak task as being more difficult, it is the task which is most readily automated.

Discussion

In deciding whether or not this approach is useful, two criteria seem appropriate. Is internal validity and experimental control sufficient to yield results which are both interpretable and statistically significant? Do these results offer insights which are useful in bridging the apparent discontinuities between human performance of laboratory and real world tasks? Although specific statistical analyses were not included in the brief synopses presented here, they are available and generally significant at the .01 level (Porter, 1987). Our confidence in the statistical significance of differences within particular experiments, is bolstered considerably by the general coherence and consistency between experiments. In answer to the second question, the results of these experiments appear to directly refute several widely-held assumptions about human performance. Although subjects can learn to perform complex tasks they may not be able to provide accurate verbal explanations of their performance. At least in this game, a particular task which was considered to be much more "difficult" (the kayak task), showed considerable resistance to intrusion from concurrent verbal side tasks as well as reduced processing time. This same "difficult" task also showed clear evidence of the development of automaticity, while a simple (but uncertain) task showed little evidence. Tasks do differ from each other, and these differences are critical in predicting human performance in situations where substance, intention and emotion are involved (i.e., real tasks).

References

Eysenck, M. W. (1984). A Handbook of Cognitive Psychology. Hillsdale, N.J.: Erlbaum.

Hitch, G. J. & Baddely, A. D. (1976). Verbal reasoning and working memory. Quarterly Journal of Experimental Psychology, 28, 603-621.

McGraw, K. O. & McCullers, J. C. (1976). Evidence of a detrimental effect of extrinsic incentives on breaking a mental set. Journal of Experimental Social Psychology, 15, 285-294.

Porter, D. B. (1987). Dual task decrements, rated difficulty and the validity of explicit knowledge. Presented at the 95th Annual Convention of the American Psychological Association at New York City, 31 August.

Wickens, C. D. (1984). Engineering Psychology and Human Performance. Columbus, OH.: Merrill.

Effect of Direction Key Reversal on Learning of Different Complexity Tasks

William C. Agisotelis and Paul K. Daly
U. S. Air Force Academy

Abstract

The effect of Stimulus Response Compatibility (SRC) on different complexity perceptual-motor tasks was investigated. Thirty freshman students aged 18 to 22 from the U. S. Air Force Academy were subjects. Score on an arcade-style computer game was measured, while direction key reversal and task complexity were manipulated. Greater SRC enhanced performance of a simple perceptual-motor task, but did not significantly improve performance of a complex task. Although full reversal caused a larger decrement in performance of the simple task, half reversal caused the greater decrement in the more cognitively-demanding equal priority game condition. Findings support a two-stage arousal theory.

People acquire habits of performing tasks in certain ways throughout life. Once learned, these habits become very hard to change. Fitts & Deininger (1954) found that when performance is affected by a nonpreferred relationship, the task becomes one where the cognitive element plays a large part. The topic of interest here is stimulus-response compatibility (SRC); Fitts and Seeger (1953) describe stimulus-response (S-R) relations as being compatible to the extent that the S-R combination comprising the task results in a high rate of information transfer. Along the same lines, Harm and Lappin (1973) concluded that S-R compatibility acted upon a translational stage of information processing, where perception is translated into action. Performance on a perceptual motor task should be at its maximum level when the task necessitates a minimum amount of information transfer; that is, when information generated by the successive stimulus events is appropriate to the set of responses that must be made in the task (Fitts et al., 1953).

The effect of differing amounts of SRC on performance of a task can also be described in terms of arousal theory. Broadbent (1971) described two arousal mechanisms which came into play: a Lower Mechanism, primarily involved in the execution of well-established decision processes, and an Upper Mechanism which monitors and alters the parameters of the Lower Mechanism to maintain the level of performance. The Lower Mechanism would operate on fairly simple tasks with high SRC, and the Upper mechanism would come into play during low SRC conditions, maintaining performance by compensating for the S-R disruption.

It has also been found that complete reversal of the display-control relationship yields better performance than do more irregular, or random, rearrangements (Gagne et al., 1950; Morin & Grant, 1955). The subject will have to think about which two of the four keys are switched, instead of following a simple rule to reverse both vertical and horizontal keys. More information transfer will take place, and performance will be lower.

The purpose of this study is to determine the effect of S-R compatibility reversal on different complexity perceptual-motor tasks. It is hypothesized that performance on all tasks will be lower on partial direction-key reversal than on complete reversal since more information transfer takes place. Also, performance of partially reversed groups on more complex tasks will be lower than on lesser complexity tasks, since Upper mechanisms are enacted, and more information transfer takes place.

Method

Subjects

30 freshman students from the United States Air Force Academy served as voluntary participants. Seven of the 30 were female, and the ages ranged from 18 to 22 years. No subjects had played the computer game before, although all had computer experience. Treatment of the subjects was in accordance with APA standards.

Apparatus

The Whale Game, developed by Porter (1987) was used as the perceptual-motor learning task. Subjects controlled the direction of movement of a whale on the screen by pressing the "2" key for up, the "S" key for down, "Q" for left, and "E" for right. There were two tasks present in each trial. The plankton task was simple but uncertain; the goal was to have the whale eat the plankton which randomly moved across the screen. The kayak task was complex but predictable; the goal was to crash the kayak into an iceberg--kayaks moved in a direct line towards the whale, and crashed when an iceberg was directly between the two. For a detailed description of the program, refer to Porter (1987). The program was modified as follows. The control version had normal direction-key pairs, as above, for all 12 trials. For the half-reversal version, the left and right directions corresponding to the horizontal keys were reversed for the last six trials. For the full-reversal version, up & down and left & right directions corresponding to both the horizontal and vertical keys were reversed for the last six trials. The game was run on Zenith 248 system computers using a BASIC system.

Procedure

Subjects were run in three groups of ten over a period of five days. The experiment was run in the morning, and lasted

approximately 45 minutes. Subjects randomly selected a computer terminal as they arrived at the testing center. They were read instructions advising them of proper use of the keyboard, goals of the tasks, and when to stop.

Subjects then ran the first six practice trials. The first trial was a plankton priority task; plankton was worth 100 points a bite and crashing or being harpooned by kayaks was worth plus or minus 10 points. The second trial was equal priority, all events were worth 50 points. The third trial was a kayak priority task, points were reversed from those of the first trial. Priority was designated by written instructions displayed on the screen before each trial. The fourth, fifth, and sixth trials were the same as the first, second, and third, respectively. After the first six trials, subjects were counterbalanced into the three reversal condition groups according to their ranked average score to counterbalance the skills across groups, and were then assigned to different terminals. The last six trials were then run in the same manner as the first six.

Results

ANOVAs for practice trials revealed no significant differences between any means. A one-way ANOVA of reversal condition means for the plankton priority revealed a significant difference between the means. A Tukey test revealed no significant difference between half-reverse and full-reverse conditions at the .05 level of significance. It was determined at the .01 level of significance that there was a significant difference between the means of control and half-reverse conditions, and between means of control and full-reverse conditions.

A separate ANOVA for the equal priority trials revealed a significant difference between means. A Tukey test was again run, and the results were the same as for the plankton priority above.

A third ANOVA was run for the kayak priority trials and revealed no significant different between the means of the three groups. A Tukey test also found no significant difference between any means at the .01 level of significance.

Graphs of number of plankton eaten vs. number of kayaks crashed (i.e., Performance Operarting Conditions) were plotted for each of the conditions (See Figure 1 below). A Mann-Whitney U test of the difference on performance efficiency between the plankton priority and equal priority trials for the half and full reversal conditions revealed at the .001 level that these curves differed significantly. The crossover between the two curves reflects this difference graphically.

Discussion

The results of this experiment show that there is a significant difference in performance when direction keys are reversed. Results showed that on plankton priority and equal

priority trials there was a significant difference in mean scores
between control and both the full and half reversal conditions,
but not between half and full reversal conditions. This suggests
that S-R compatibility has an effect on performance of a
perceptual-motor task, but the degree of reversal has little
effect on simple but uncertain, or combined tasks. For the kayak
task, no significant difference between means of any reversal
conditions was found, which might suggest that S-R compatibility
has little effect on this more difficult task. The relative
resilience of the kayak task to cognitively demanding side tasks
or manipulations is consistent with Porter's (1987) findings.
Our findings failed to support our hypothesis that performance
would be lower with half-reversed groups than full-reversed
groups. However, looking at the graph (Fig. 1), there is a
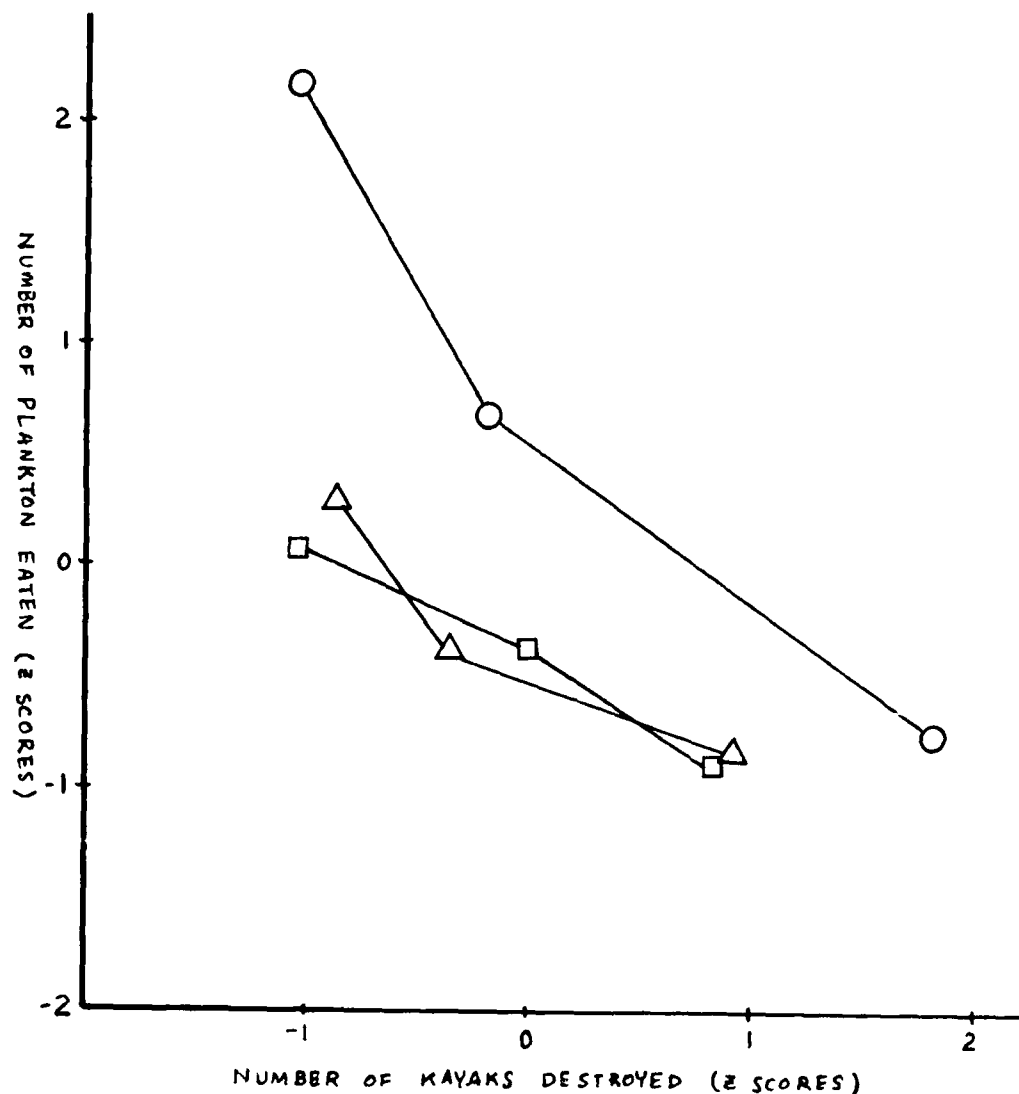
Figure 1. Plankton eaten vs. kayaks crashed.

crossover; scores on the equal priority task are higher for fullrev than halfrev. This suggests that there is a greater cost to combining tasks in half-reversal trials, and that there is good time-sharing between tasks in full-reversal trials. The Mann-Whitney U test showed that this crossover was in fact significant.

This crossover can be explained using Broadbent's two-stage arousal theory. Half reversal seems to be more difficult, and seems to involve the Upper mechanism. The subject has to remember two things: which keys are reversed, and how to then move in the appropriate direction. The full reversal condition, on the other hand, might involve the Lower mechanism; the subject uses a simple transformational algorithm—just to go the reverse of the previous trials. The crossover occurs in the equal priority condition since the subject now has to deal with both tasks instead of concentrating on one. The cognitive element now plays a big part, and the Upper and Lower arousal mechanisms come into play, more so than in the simple plankton priority trial.

It has been shown that S-R incompatibility does have an effect on perceptual-motor performance tasks, but the influence of degree of compatibility on varying complexity tasks is not as clear. It appears that different cognitive mechanisms are at work in different SRC designs, but more research needs to be done in this area.

References

Broadbent, D. E. (1971). Decision and stress. New York: Academic Press.

Fitts, P. M., & Deininger, R. L. (1954). S-r compatibility: Correspondence among paired elements within stimulus and response codes. Journal of Experimental Psychology, 88, 483-492.

Fitts, P. M., & Seeger, C. M. (1953). S r compatibility: Spatial characteristics of stimulus and response codes. Journal of Experimental Psychology, 46, 199-210.

Gagne, R. M., Baker, K. E., & Foster, H. (1950). On the relation between similarity and transfer of training in the learning of discriminative motor tasks. Psychological Review, 57, 67-79.

Harm, O. J., & Lappin, J. S. (1973). Problem compatibility speed and accuracy. Journal of Experimental Psychology, 100, 416-418.

Morin, R. E., & Grant, D. A. (1955). Learning and performance on a key-pressing task as function of the degree of spatial stimulus-response correspondence. Journal of Experimental Psychology, 88, 39-47.

Porter, D. B. (1987, August). Dual task decrements, rated difficulty and the validity of explicit knowledge. Paper presented at the 95th Annual Convention of the American Psychological Association, New York.

Effects of Verbalization on Single and Dual Task Performance

David M. Koch
Troy J. Ziegler
David B. Porter
U.S. Air Force Academy

ABSTRACT

The purpose of this experiment was to determine the effects of concurrent articulation on subjects' performance of a complex computer game. Twenty subjects were instructed to continuously verbalize the direction of movements of the primary game character. This prevented them from talking to themselves while they played the game. The twenty subjects in the control group played the game without concurrent verbalization. The game consisted of two substantively different tasks combined under three different priority instructions. There were no significant differences between experimental and control group performance on either single task, however, our results suggest concurrent verbalization significantly shifted subjects' bias and improved performance efficiency during equal priority trials.

People often talk to themselves while they perform a task. Sometimes they talk as if instructing themselves, other times they criticize themselves, and other times they praise themselves. It is difficult to determine what effect this cognitive kibitzing has on performance and learning. It might seem that talking oneself through a task would aid in memorizing what had to be done, but it also seems that talking to oneself might interfere with a largely non-verbal task.

Our experiment set out to test several of the assertions made by tennis pro Tim Gallwey in The Inner Game of Tennis. Gallwey describes separate cognitive entities: Self 1 and Self 2. Self 1 is the conscious verbal mechanism and Self 2 is the unconscious controller of physical actions. "If Self 1 is always yelling at and criticizing Self 2, performance is bound to suffer. Indeed, peak performance, usually occurs when the verbal self completely stands aside"(Gallwey, 1977, p. 27). By having our subjects verbalize the direction of movement of the video character they controlled, their articulatory loops could not be used for self instruction. This should cause their performance to increase since they will be more focused on the task and not free to criticize themselves.

METHOD

The video game used was called simply the Whale Game
(Porter, 1987). The screen displayed an ocean with four clusters
of icebergs. Subjects controlled the direction of movement of a
whale by using four keyboard keys. The whale could move up, down,
left and right. It could not, however, move diagonally nor cross
screen boundaries. Points were scored for either having the whale
eat plankton or wreck kayaks. The points scored for each event was
changed by the priority instructions provided prior to each game
trial. There were three distinct priority conditions: plankton,
equal and kayak. Priority was changed on each successive trial and
counterbalanced both within and between groups.

The plankton task was "simple but uncertain." (Porter, 1987)
A mass of plankton began each trial in the middle of the screen.
The player had to guide the whale into the plankton to score
points. The plankton moved randomly and its direction of travel
could not be predicted by the subject. This task was very similar
to a classic two dimensional pursuit tracking task except discrete
(0-order) digital inputs were required.

In contrast, the kayak task was "complex but predictable."
(Porter, 1987) Kayaks were generated at five places along the
border of the screen at fixed time intervals. The kayaks always
moved straight toward the whale. If the whale was hit by a kayak,
points were deducted from the score. If the whale was maneuvered
so that a kayak hit an iceberg, the kayak was destroyed and points
were added to the score. Since each kayak moved directly toward
the whale, the whale's position simultaneously influenced the
paths of all kayaks on the screen.

There were three game priorities determined by the point
value assigned to the two tasks. In one priority condition (P),
the plankton was worth 100 points if eaten, and the kayaks were
worth 10 points if smashed into icebergs (ten points were deducted
for each kayak that harpooned the whale). In another priority
condition (E), the plankton was worth 50 points and the kayaks
were worth plus or minus 50 points. In the third priority
condition (K), the plankton was worth 10 points and the kayaks
were worth plus or minus 100 points. Priority was designated by
instructions displayed on the screen before each trial.

The subjects all completed 3 practice games (one of each
priority). They then completed 9 games (three of each priority).
The games were rotated in either a P,E,K order or a K,E,P order to
counterbalance the effects of practice on the two tasks. To insure
subjects had knowledge of their own performance, they each
recorded the number of kayaks destroyed and the tons of plankton
eaten after each two and one half minute trial.

There were a total of forty subjects. All subjects were
fourthclass cadets at the USAF Academy between 18 and 21
years of age. Twenty subjects were instructed to continuously

262

verbalize the direction of movement of their whale while playing
the game. This manipulation occupied their articulatory loops, and
also served to focus their attention on the whale. Subjects in
the control group were not given any additional instruction and
were free to articulate as they pleased. Subjects completed the
experiment in groups of 6 to 10 with other subjects in the same
group (i.e., experimental or control).

RESULTS

A review of subjects' performance suggested that 4 of the 40
subjects did not followed the priority instructions(i.e., had not
shifted their concentration from one task to the other). These
four subjects were replaced.

The average tons of plankton eaten for the three plankton
priority trials for the concurrent verbalization
subjects was compared to the corresponding average for the
control subjects. There was not a significant difference between
these scores $(t(38)=.103, p>.05)$. The average number of kayaks
destroyed (KD) during the kayak priority trials for the
experimental group was also compared to the average for the
control group. Again, no significant difference was found $(t(38)=
.26, p>.05)$. Although the group with the verbal sidetask performed
slightly better than the control group on both single tasks,
neither difference was significant.

The equal priority task (where each task was worth fifty
points) had more interesting results. Separate measures of
efficiency and bias were computed. The bias calculations were
computed using the formula P-2K (Plankton - 2 x Kayaks). The
control group's average equal priority task performance was found
to be significantly biased toward the kayak task compared to the
performance of the group with the sidetask $(t(38) = 2.42, p <
.02)$. The average performance efficiency (P + 2K) of the group
with the verbal side task was also found to be significantly
better than the control groups performance during non-verbal equal
priority tasks $(t(38) = 2.07, p<.05)$.

One way to represent these results graphically is through
Performance Operating Characteristics (POC's) (Wickens, 1984).
Subjects' average performance on the plankton task is shown along
the vertical axis and their performance on the kayak task along
the horizontal axis. Separate POC's for the control and
experimental groups are formed by connecting average performances
of both tasks during each of the three priority conditions. The
significant differences in both the bias and efficiency of the two
groups is readily apparent.

DISCUSSION

These results are interesting for several reasons. The lack
of interference with the plankton task caused by the verbal side

figure 1.

Effects of Verbalization on Single and Dual Task Performance.

experimental ————

control — — — —

* efficiency = P + 2K

** bias = P - 2K

PLANKTON EATING

KAYAK CRASHING
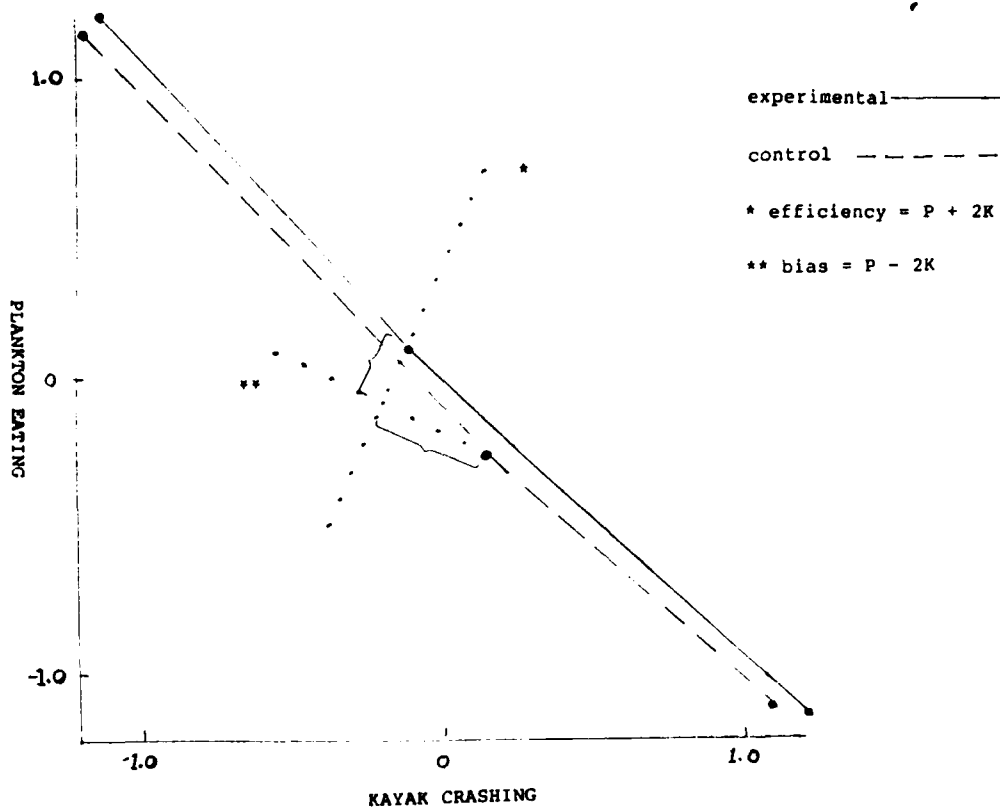
task appears to contradict several earlier findings (Porter, 1987). Other concurrent verbal side tasks had caused significant decrements to this simple but uncertain task. Our particular verbal side task, however, not only occupied the articulatory loop, it also kept subjects' attention focused rather narrowly on the movement of the whale. Theoretically this should have facilitated performance of the relatively simple plankton task but inhibited performance of the complex kayak task (Easterbrook,1959). Thus the two effects of our side task would have simply cancelled each other out. Despite these onjectures, the only clear result is that this particular verbal side task showed no significant effect on either of the game tasks when they were performed in their respective "high priority" conditions. These negative results served to make the significant differences in the performance of the two groups during equal priority trials even more interesting.

Occupying subjects articulatory loops with concurrent verbal side task actually improved their performance (efficiency) during what is arguably the most difficult and demanding priority condition (i.e., combining the two tasks). This finding clearly contradicts predictions of most traditional "resource models" (Wickens, 1984). The significant shift in bias away from the kayak task for the experimental group suggests a partial explanation.

Subjects in the control group appeared to be inappropriately biased toward the more complex and subjectively more difficult kayak task. Although this shift in bias did result in a significant decrease in their performance of the plankton task it did not result in increased performance on the kayak task. This is consistent with Gallwey's suggestion that verbal Self 1 may actually interfere with the natural performance of the non-verbal Self 2. It is also consistent with Porter's (1987) earlier findings that resource demanding side tasks decreased performance on the plankton task but the provision of additional processing resources did not significantly improve performance of the kayak task. These findings also lend support to other work on implicit learning. As Reber and Kassin (1980) suggest:

"...complex structures such as those underlying language, socialization, perception and sophisticated games are acquired implicitly and unconsciously; such knowledge is memorially en:oded in an abstract representational system... the acquisition process itself contains at its core an induction process whereby relations among parts of the stimulus environment are mapped in an order that corresponds roughly with the ecological salience of these relations." (p.493)

In other words, when individuals don't have access to accurate verbal information about the task, interfering with their capacity for subvocal articulation may actually improve their performance (i.e., Gallwey may be correct).

REFERENCES

Easterbrook, J.A. (1959). The effect of emotion on cur utilization and organization of behavior. Psychological Review, 66, 183-201.

Gallwey, W. T. (1974). The Inner game of Tennis. New York: Random House.

Porter, D.B. (1987). Dual Task Decrements, Rated Difficulty and the Validity of Explicit Knowledge. USAFA: Dept. of Behavioral Science and Leadership.

Reber, A.S. and Kassin, S.M. (1980). On the relationship between implicit and explicit modes in the learning of a complex rule structure. Journal of experimental Psychology. Human Learning and Memory Vol 6. No. 5, 492-502.

Wickens, C.D. (1984). Engineering Psychology and Human Performance. Columbus, Ohio: Merrill Publishing.

The Role of Consistency in the Development of Automatic Processing

Arthur F. Kramer, David L. Strayer and Yili Liu

Department of Psychology
University of Illinois

Abstract

This study examined the role of consistency in the
development of automatic processing. Subjects performed a memory
search task with either consistent (CM) or varied (VM) mapping of
rules to responses. In both of these conditions the mapping of
rule components to responses was varied over trials. Measures of
mean reaction time and A´ indicated improved performance in both
CM and VM rule to response mapping conditions. However,
performance in the CM conditions was superior to that obtained in
the VM conditions. The results suggest that when higher-order
consistency is present in a task consistent mapping of stimulus
elements to responses is not necessary for the development of
automatic processing.

In recent years a number of models have been proposed to account for the
quantitative and qualitative changes in performance that occur during the
acquisition of complex skills. A common element in these models is the
suggestion that "consistency" among task components is necessary for the
development of highly skilled behaviors (LaBerge, 1981; Logan, in press;
Posner & Snyder, 1975). In one such model of skill acquisition, Schneider
and Shiffrin (1977) proposed that two different modes of processing underlie
performance. Automatic processing which develops as a result of extensive
practice with consistent stimulus-response relations is characterized as
fast, inflexible, difficult to supress once learned, and not limited by
short term memory capacity or attention. Controlled processing occurs in
novel situations or in situations in which stimulus-response relations are
inconsistent over time. This mode of processing is characterized as slow,
serial and capacity limited.

In the present study we examine the role of consistency in the
development of automatic processing. Within the visual/memory search
literature consistency has been defined as the mapping between stimuli and
responses. In a consistent mapping condition (CM) targets are always
selected from one set of items (e.g. letters A to M) while distractors are
selected from another set of items (e.g. letters N to Z). Thus, the mapping
of stimuli to responses does not vary over trials in the CM conditions.
However, in varied mapping (VM) conditions targets and distractors are
selected from the same set of items (e.g. letters A to Z). Targets and
distractors exchange roles over trials in the VM conditions. Automatic
processing develops in the CM but not in the VM conditions.

Although numerous laboratory studies have demonstrated the value of
stimulus-response consistency in the development of automatic processing,
the examination of a number of "real-world" tasks suggests other types of
consistency might also be important. For example, medical diagnosis
requires that diagnosticians associate symptoms with disease states.
However, unlike simple S-R consistency the same symptom may be associated
with a number of different diseases. Thus, in the task of medical diagnosis
consistency occurs at the level of conjunctions of symptoms rather than at a
symptom to disease (e.g. S-R) level.
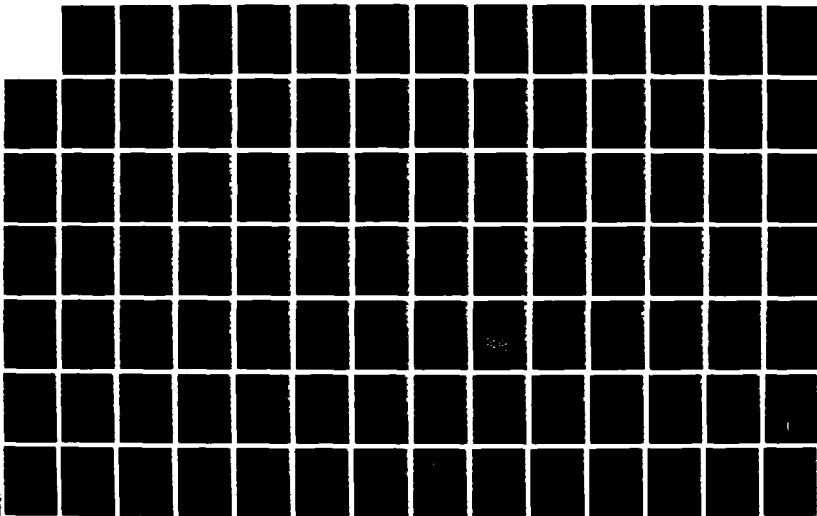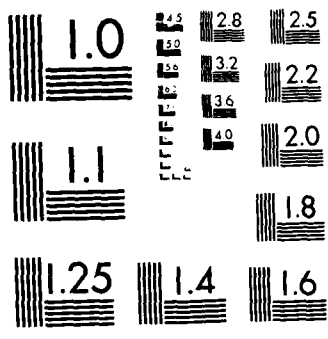
Duncan (1986) in a critical review of automaticity has argued that, "... the important question is not whether mapping is consistent, but at what level of stimulus description it is consistent, and at what level consistency affects learning" (pg. 283). In a series of recent studies Durso et al. (1987) and Fisk et al. (in press) have shown that automaticity can be achieved even when S-R consistency is not maintained. In their studies, subjects responded to either the highest or lowest digit in a display. In this case although the ordinal relations among numbers are consistent the stimuli are not consistently mapped to responses (e.g. 8 could be the highest number on one trial, while 9 was the highest number on another trial). Both sets of investigators found decreasing RT´s and reduced display size effects with practice.

In the present study we further explore the role of consistency in the development of automatic processing by requiring subjects to perform a memory search task with consistent or varied rule to response mapping. However, in both CM and VM conditions the mapping of rule components to responses is varied. Thus, the main question addressed in this study is whether consistency at the rule level will be sufficient for the development of automaticity when stimulus level (e.g. S-R) consistency is absent.

Method

Subjects

Eight right handed persons, aged 20 to 28 years were recruited from the student population at the University of Illinois. All of the subjects had normal or corrected to normal vision.

Procedure

Subjects performed a variant of the Sternberg (1966) memory search task. In each block of trials subjects (a) memorized either one, two or three rules and (b) compared the 30 probe stimuli that followed the presentation of the memory set to the rules. A CRT response was used to indicate whether each of the probes matched a memory set rule. Response hand was counterbalanced across subjects. Memory sets were presented for 6 secs. Probes were presented for 200 msec. ISIs were 4 sec.

The probe stimuli consisted of four concentric rings and two numbers presented in the center of the display. Numbers could appear anywhere within the four rings. The memory set rules indicated the relationship between the two numbers and the rings in which the numbers would be presented. For example, one rule indicated that the numbers would be presented one ring apart and the inner digit would be less than the outer digit by two. Another rule indicated that the numbers would occur two rings apart and the inner digit would be greater than the outer digit by three.

Sixteen separate rules were used in the experiment. Eight rules were assigned to the VM condition. The other eight rules were assigned to the CM conditions (four target and four distractor rules). For half of the subjects the numbers that were used in the rings for sessions one through twelve were selected from 0 to 4. These subjects were transfered to the numbers 5 to 9 in blocks thirteen through eighteen. The other four subjects started with the numbers 5 to 9 and transferred to 0 to 4. It is important to note that the numbers and the position of the numbers in the rings were variably mapped across trials in the CM and VM conditions. However, the rules were either consistently (CM condition) or variably mapped across

trials (VM condition).

Each of the subjects participated in 18 experimental sessions. Sessions lasted approximately two hours. During each session subjects performed 36 blocks of 30 trials (1080 trials per session). Subjects were presented with RT and accuracy feedback after each block of 30 trials.

Design

The experimental design was a within-subject four-way factorial. The factors were mapping condition (CM and VM), memory set size (1, 2 and 3), response type (target and distractor) and phase (training and transfer). Subjects participated in 12 sessions in the training phase and six sessions after transfer to a different set of numbers.

Results

Figure 1 presents the mean correct RTs as a function of session, mapping condition and memory set size. We will address first the data from the training condition (sessions 13 to 18). All reported effects are significant at $p<.05$.

The significant main effect for session ($F(1,7)=226$) confirmed the reduced RTs with practice that can be seen in figure 1. Main effects were also obtained for memory set size ($F(2,14)=96$) and mapping ($F(1,7)=34$), with longer RTs for the larger memory set sizes and VM conditions. The significant three-way interaction among memory set size, session and mapping condition ($F(2,14)=4$) suggests that the decrease in the memory set size effect with practice was larger in the CM than the VM conditions. This difference in memory set size slope with practice is one criterion used to distinguish between automatic and controlled processing (Schneider & Shiffrin, 1977).
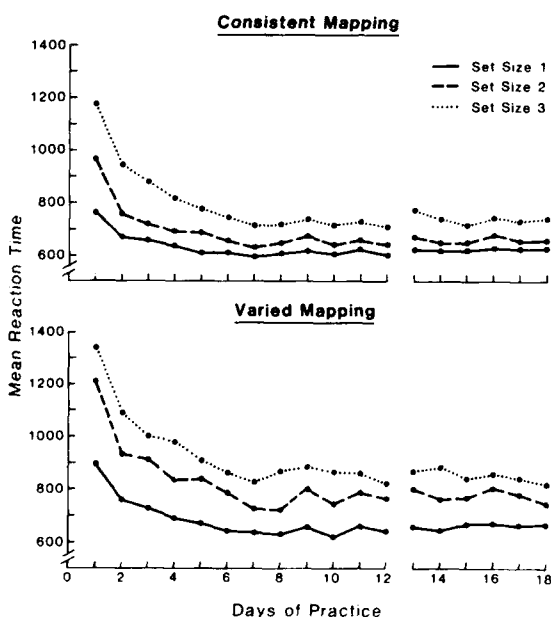
Figure 1 Mean correct RTs as a function of session, mapping condition and memory set size for the target probes.
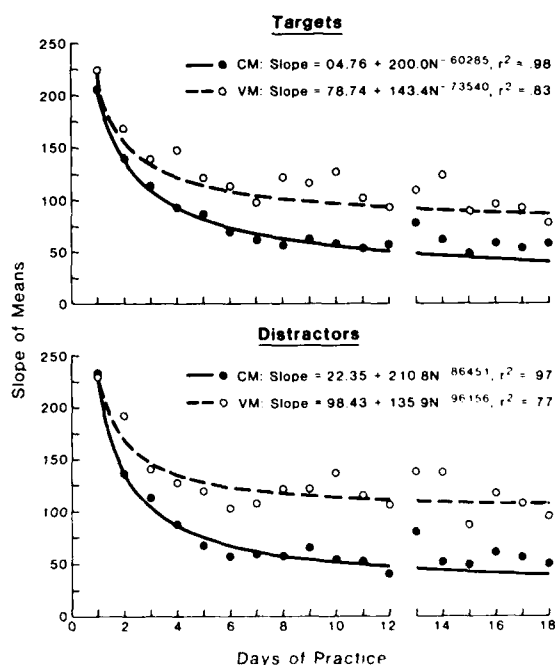
Figure 2 Mean RT memory set slopes as a function of mapping condition, session and response type.

268

The extent of the decrease in memory set slope for the CM and VM conditions with practice can be seen in figure 2. Although the slopes for CM and VM conditions began at essentially the same point in session 1, the CM slopes asymptoted at a lower value than the VM slopes after 12 sessions of practice. The reduction in slopes for the CM and VM target conditions was 73% (205 to 55 msec) and 60% (222 to 90 msec), respectively. The reductions for the CM and VM distractor conditions was 83% (233 to 40 msec) and 54% (229 to 106). The power law functions that have been fit to slopes indicate two additional differences between CM and VM conditions. First the R squared values suggest a better fit for the CM than the VM conditions. Second, the intercept constants (e.g. 4.76 for CM targets, 78.74 for VM targets) indicate that the asymptotic slopes will be substantially smaller in the CM than in the VM conditions.

Figure 3 presents the nonparametric A´ measures of detection sensitivity as a function of session, memory set size and mapping condition. A´ ranges from .5 for chance accuracy to 1.0 for perfect detection accuracy. A´ was quite high in all conditions. However, A´ was larger for the CM than the VM conditions ($F(1,7)=27$), but this difference decreased with practice ($F(2,14)=8$). The difference in A´ as a function of memory set size was larger for the VM than the CM conditions ($F(2,14)=6$).

The transfer condition (sessions 13 through 18) was conducted to determine whether the level of performance attained in the CM conditions could be attributed to the automatic processing of (a) the rules or (b) the conjunctions of features of the multidimensional stimuli or (c) some combination of a and b. It is conceivable that subjects may have learned the 78 conjunctions of numbers and positional arrangements that satisfied the 16 rules rather than processing the probes on the basis of the rules. Thus, subjects may have generated categories for each of the rules with the category exemplars represented by the combination of features that satisfy each rule. If this strategy were adopted it would be expected that transfer to a new set of numbers would dramatically decrease task performance. If, on the other hand, subjects had been processing the stimuli on the basis of the rules the transfer to a different set of stimulus features (e.g. numbers) should have a minimal impact on performance.
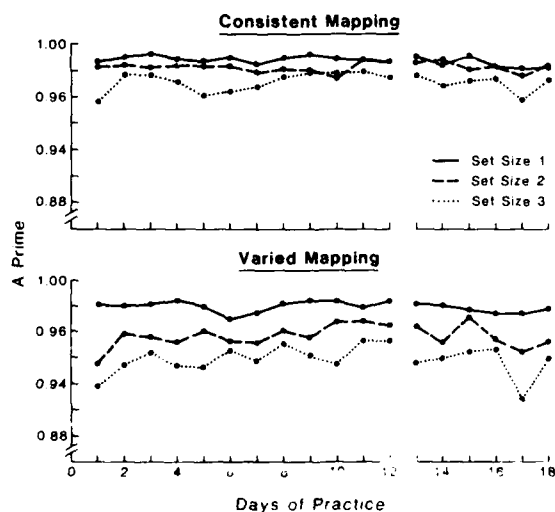
Figure 3 Mean A´ values as a function of session, mapping condition and memory set size.

269

In order to address the issue of transfer to a different set of stimulus features, repeated measures ANOVAs were performed for the RT and A´ variables. The factors for the RT ANOVA were sessions (12 through 18), mapping (CM & VM), memory set size (1, 2 & 3) and response type (targets & distractors). The A´ ANOVA included the session, mapping and set size factors. For RT, the only significant effect involving the session factor was a set size by session interaction $(F(12,84)=3)$. RT was elevated in session 13 for the largest set size. Neither the main effect of session nor its interaction with other factors reached significance for the A´ measure.

Consistent with the findings of Druso et al. (1987) and Fisk et al. (in press) the results obtained in the present study suggest that the consistent mapping of stimuli to responses is not critical for the development of automatic processing given that subjects can capitalize on higher-order consistencies in a task. The transfer results suggest that subjects are capable of automatic processing at the rule level, at least in situations in which the alternative is to learn a large number of category exemplars. Finally, the learning curves (see figures 1 & 2) indicated that significant gains in performance occurred in both CM and VM rule to response mapping conditions. This may seem at odds with the typical finding of an invariant slope with practice in VM search conditions. However, in the present experiment the rules did not vary in the CM and VM conditions, only the mapping of the rules to responses. Thus, even in the VM condition there was consistent mapping within rules.

Acknowledgment

References

Duncan, J. (1986). Consistent and varied training in the theory of automatic and controlled information processing. Cognition, 23, 279-284.

Durso, F.T., Cooke, N.M., Breen, T.J. & Schvaneveldt, R.W. (1987). Is consistent mapping necessary for high-speed search. Journal of Experimental Psychology: Learning, Memory and Cognition, 13, 223-229.

Fisk, A.D., Oransky, N.A. & Skedsvold, P.R. (in press). Examination of the role of higher order consistency in skill development. Human Factors.

LaBerge, D. (1981). Acquisition of automatic processing in perceptual and associative learning. In P.M.A. Rabbitt & S. Dornic (Eds.), Attention and Performance V. New York: Academic Press.

Logan, G. (in press). Toward an instance theory of automatization. Psychological Review.

Posner, M. & Snyder, C. (1975). Attention and cognitive control. In R. Solso (Ed.), Information Processing and Cognition: The Loyola Symposium. Potomac, Maryland: Erlbaum.

Schneider, W. & Shiffrin, R. (1977). Controlled and automatic human information processing: I. Detection, search and attention. Psychological Review, 84, 1-66.

Sternberg, S. (1966). High speed scanning in human memory. Science, 153, 652-654.

Adult Age Differences in the Development of Automaticity:
A Psychophysiological Assessment

David L. Strayer, William J. Gehring, Arthur F. Kramer, and Gregory A. Miller

University of Illinois at Urbana-Champaign

Abstract

This study provides a fine-grained analysis of the age-related differences in the acquisition of automatic processing. Following consistent mapping (CM) training, young subjects develop automatic processing. While elderly subjects improve under CM training conditions, this is not due to the development of automatic processing, but to a reduction in response-related processing. Older subjects adopted a more conservative response bias than the young, which may interfere with the development of automaticity.

One of the most ubiquitous age-related changes in performance is the slowing of behavior with senescence. However, automatic skills which are acquired in young adulthood appear to be resistant to these decrements (Casey et al., 1987). Several investigators have sought to determine if the elderly can develop automatic processing skills and thereby eliminate the age-related decrements in performance. Studies which have addressed this issue have found that the elderly generally do not attain the same level of asymptotic performance in tasks that require automatic processing, nor do they improve at the same rate as young adults (e.g., Madden & Nebes, 1980; Nobel et al., 1964; Plude et al., 1983).

The purpose of this study is to provide a fine-grained analysis of the age-related differences in the development of automaticity. We will employ the converging methodologies of additive factors logic and the P300 component of the Event-Related Brain Potential (ERP) to localize the changes in information processing with practice and S-R mapping. Several lines of converging evidence suggest that the latency of the P300 component is sensitive to stimulus evaluation processes, but relatively insensitive to response-related processes (e.g., Magliero et al., 1984). Since a major portion of the slowing with age has been localized to response-related processing (Strayer et al. 1987), we seek to determine the extent to which automatic processing can bypass this information processing bottleneck. We further ask whether the stimulus evaluation processes can be automated in the elderly.

Method

Subjects

Eight young (mean age = 20.6, sd=1.5) and eight elderly (mean age = 73.1, sd=6.7) subjects participated in the experiment. All subjects had normal or corrected-to-normal vision and were in good health. Subjects were paid for their participation.

Procedure

Subjects performed a variant of the Sternberg memory search task (1966) under both consistent mapping (CM) and varied mapping (VM) conditions. Each block of trials consisted of a memory set presented for 3 seconds, followed by 30 probe trials. On each trial two probe letters were simultaneously presented within 1.5 degrees of visual angle. Subjects were instructed to press one

button if either of the probes was a member of the memory set and press another button if neither probe was a member of the memory set. The experiment was conducted in four sessions (days), resulting in 5760 CM trials and 5760 VM trials.

ERP Recordings

In sessions one and four, EEG was recorded from Fz, Cz, and Pz sites. EEG and EOG were sampled every 10 msec for 1300 msec, beginning 100 msec prior to stimulus onset. EOG artifacts were corrected off-line. Single trial estimates of P300 latency were derived using a peak picking algorithm at the Pz electrode within a window from 300 to 1150 msec. Average ERPs were generated for each experimental condition. Each subject contributed a maximum of 360 trials to each average.

Results

Our analyses will focus on the first and last sessions of practice. The design is a 2 (age: young vs old) X 2 (session: first vs last) X 2 (mapping: CM vs VM) X 2 (set size: 2 vs 4) X 2 (response type: target vs distractor) split-plot factorial. The results will be hierarchically organized. Within each dependent measure we will examine the effects of practice and S-R mapping, followed by an analysis of age-related differences in these effects. For all analyses, a significance level of .05 is adopted.

Reaction Time

Table 1 presents mean reaction time to targets as a function of age, session, mapping, and set size. The linear regression slopes for memory set size are also presented.

Table 1

S-R Mapping: Memory Set Size:	CM			VM		
	2	4	Slope	2	4	Slope
Young Session 1	469	533	31.9	494	630	68.2
Session 4	414	434	9.9	454	572	58.8
Old Session 1	652	737	42.8	685	804	59.2
Session 4	558	597	19.2	619	752	66.6

Reaction time decreased with practice $F(1,14)=19.1$, was shorter for CM than for VM conditions $F(1,14)=114.9$, and increased as a function of set size at a greater rate in VM conditions than in CM conditions $F(1,14)=61.0$. The difference between CM and VM conditions increased from session 1 to session 4, $F(1,14)=17.0$, as did the CM and VM difference as a function of memory set size, $F(1,14)=10.3$. In addition, reaction time was shorter for targets than for distractors, $F(1,14)=113.9$.

Elderly responded more slowly than young, $F(1,14)=19.5$. Both groups improved with practice. The absolute level of reduction in the memory slope for CM conditions was equivalent; however, the ratio of CM slopes in session 1 vs session 4 revealed a greater proportional reduction for the young (3.22) than for the elderly (2.23). Furthermore, the ratio of VM to CM slopes in session 4 was larger for the young (5.94) than for the elderly (3.47). This reflects differences in asymptotic levels of performance. The session 4 CM slopes of the elderly were twice the slopes of the young.

Detection Sensitivity

The non-parametric measure A' was adopted to determine changes in detection sensitivity. A' ranges from .50 for chance accuracy to 1.0 for perfect detection accuracy. A' was quite high throughout the experiment, ranging from .86 to .99. A' decreased as a function of set size, $F(1,14)=94.9$, and was larger for CM than VM conditions, $F(1,14)=100.0$; however, the decrease in A' as a function of set size was larger for VM than CM conditions, $F(1,14)=41.7$. A' also increased from session 1 to session 4, $F(1,14)=30.3$, and this was more evident for VM than CM conditions, $F(1,14)=4.45$. This interaction is probably due to a ceiling effect for CM conditions, since initial A' values were quite high.

A' increased from session 1 to session 4 more for the elderly than for the young, $F(1,14)=7.7$, reflecting poorer detection sensitivity for the elderly early in training and greater detection sensitivity for the elderly following training. Furthermore, elderly were more affected by S-R mapping, $F(1,14)=6.5$, and by set size, $F(1,14)=7.3$, than the young. These latter effects are heavily influenced by VM performance in session 1, where memory set size produced its greatest effect on the elderly.

Response Bias

The non-parametric measure B'' was used to assess subjects' response bias. Larger values of B'' reflect a more conservative response bias. In session 1, subjects adopted a more conservative response bias for CM than VM conditions; however, in session 4 the subjects responded more conservatively in VM conditions, $F(1,14)=15.5$. In addition, B'' decreased with increasing memory set size for VM conditions, but not for CM conditions, $F(1,14)=12.9$, indicating that subjects adopted a more risky response strategy to compensate for the more difficult VM condition.

Elderly responded more conservatively than young, $F(1,14)=6.2$. Furthermore, young subjects became less conservative with practice, but elderly subjects became more conservative following practice, $F(1,14)=4.53$. Elderly also tended to respond more conservatively in CM conditions than VM conditions, while young subjects tended to respond more conservatively in VM conditions than in CM conditions, $F(1,14)=6.6$. This was coupled with a tendency of the elderly to become more conservative as memory set size increased in CM conditions, $F(1,14)=9.2$.

P300 Latency

Table 2 presents mean P300 latency to targets as a function of age, session, mapping, and set size. The linear regression slopes for memory set size are also presented.

Table 2

S-R Mapping:		CM			VM		
Memory Set Size:		2	4	Slope	2	4	Slope
Young	Session 1	590	654	32.1	614	697	41.5
	Session 4	610	621	5.6	611	707	47.8
Old	Session 1	708	738	14.7	708	754	23.0
	Session 4	631	654	11.6	678	738	29.6

P300 latency increased with set size, $F(1,14)=80.1$; however, the effect was greater for VM than CM conditions, $F(1,14)=13.2$, particularly in session 4, $F(1,14)=6.1$. P300 latency was shorter for CM than VM conditions, $F(1,14)=23.8$, and this was more pronounced in session 4, $F(1,14)=16.1$. In addition, P300 latency was shorter for targets than distractors, $F(1,14)=93.7$.

P300 latency was shorter for young than elderly, $F(1,14)=8.5$. However, the effect of memory set size was greater for the young than for the elderly, $F(1,14)=8.3$. Furthermore, there were differential effects of mapping, response type, and session for young and elderly, $F(1,14)=8.5$, and mapping and set size, $F(1,14)=5.0$. For the young, the effect of memory set size produced equivalent effects on P300 latency for CM and VM conditions in session 1, but in session 4 the memory set size effect was substantially reduced in the CM condition. In contrast, the effects of memory set size were relatively constant across session for both CM and VM conditions for the elderly. The improvement with practice can be illustrated by comparing the ratio of the CM slopes in session 1 and 4 for the two age groups. The CM session1/session4 ratio for young was 5.73 and 1.27 for elderly.

RT/P300 Ratio

A single-trial ratio of RT to P300 latency was calculated to determine the proportion of stimulus evaluation accomplished at the moment of response in each condition. A ratio of 1.0 indicates that the RT response and peak of the P300 co-occurred. Ratios less than 1.0 indicate that the response preceeded P300 latency and ratios greater than 1.0 indicate that the response followed P300 latency. Previous research suggested that a large portion of the age-related slowing is due to response-related processing (Strayer et al., 1987). This analysis was conducted to determine if post-stimulus evaluation processing was reduced following consistent practice in the elderly.

The RT/P300 ratio decreased with practice, $F(1,14)=18.8$, and increased with set size, $F(1,14)=53.2$. This latter effect was more evident in session 1, $F(1,14)=6.7$. Further, the RT/P300 ratio was larger for VM than CM conditions, $F(1,14)=41.0$, and this was more pronounced in set size 4, $F(1,14)=20.5$.

The RT/P300 ratio was larger for elderly than young, $F(1,14)=10.1$. The average RT/P300 ratio was 0.89 for the young and 1.12 for the elderly. This implies that elderly engaged in more post-stimulus evaluation processing prior to their response than the young. Age did not enter into any interactions, suggesting that the age differences in post-stimulus evaluation were not modified by practice or S-R mapping.

Discussion

Both young and elderly improved with practice. This improvement was observed as decreases in reaction time and increases in response accuracy. However, elderly responded more slowly and more accurately than young following practice. This suggests that the elderly were trading response speed for accuracy, adopting a more conservative response bias than the young.

Reductions in the reaction time memory set size slope were apparent for both age groups, but the young improved more rapidly and achieved a lower asymptote. Reductions in the P300 latency memory set size slopes were apparent only for the young. This suggests that the stimulus evaluation processes become automated for the young, but not for the elderly. The improvement in reaction time performance for the elderly was attributed to a reduction in response-related processing which was apparent as a trimming of long latency

274

responses. This pattern of data suggests that the improvements in performance of the elderly are not the result of automatic processing, but rather are due to a reduction in post stimulus evaluation processing.

One possible interpretation for why the elderly do not acquire automatic processing may be their conservative response bias. It has been suggested that such a conservative strategy interferes with the development of automaticity (Shiffrin et al., 1984). One prediction from this interpretation is that if elderly subjects adopted a less conservative response bias, then their acquisition rates and asymptotes should be similar to young. Further, if young subjects adopt a more conservative response bias, then the rate of improvement and the asymptote should be similar to that of the elderly.

Thus age differences in the development of automaticity appear to be the result of strategic changes in information processing.

References

Casey, B.J., Rogers, W.A., & Fisk, A.D. (1987). Cognitive skill acquisition: A developmental approach. Proceedings of the 31st Annual Meeting of the Human Factors Society. New York, New York.

Madden, D.J., & Nebes, R.D. (1980). Aging and the development of automaticity in visual search. Developmental Psychology, 16, 377-384.

Magliero, A., Bashore, T.R., Coles, M.G.H., & Donchin, E. (1984). On the dependence of P300 latency on stimulus evaluation processes. Psychophysiology, 21, 171-186.

Nobel, C.E., Baker, B., & Jones, T.A. (1964). Age and sex parameters in psychomotor learning. Perceptual and Motor Skills, 19,935-945.

Plude, D.J., Kaye, D.B., Hoyer, W.J., Post, T.A., Saynisch, M.J., & Hahn, M.V. (1983). Aging and visual search under consistent and varied mapping. Developmental Psychology, 19, 508-512.

Shiffrin, R.M., Dumais, S.T., & Schneider, W. (1984). Characteristics of automatism. In J. Long and A. Baddeley (Eds.), Attention and Performance IX. Erlbaum: Hillsdale, New Jersey.

Sternberg, S. (1966). High speed scanning in human memory. Science, 153, 652-654.

Strayer, D.L., Wickens, C.D., & Braune, R. (1987). Adult age differences in speed and capacity of information processing: 2. An electrophysiological approach. Psychology and Aging, 2, 99-110.

The Effects of Level of Processing and Incentive on Recall in a Incidental Learning Environment

Edward M. Cavello,
Lorelei Faber,
Scott L. Smith,
U.S. Air Force Academy

Abstract

The purpose of this study was to examine the effects of levels of processing and incentive on recall in an incidental learning environment. Four groups were used to examine the effects of the two independent variables on the quantity and quality of words recalled from an auditory list of forty adjectives. Level of processing and incentive had additive effects which supported earlier studies that semantically processed information will be recalled better and that incentive will also improve recall. This also implies that these two factors influence different aspects of the human information processing system.

This study examines how words are processed, stored, and subsequently recalled. The hypothesis being tested is that both deeper levels of processing and incentives will increase recall in an incidental learning situation.

Craik and Lockhart (1972) developed Levels-Of-Processing Theory. Their first major assumption was that memory should not be regarded in isolation and that memory traces are formed as by-products of attentional and perceptual processes. Their most important theoretical assumption was that memory is a function of depth of analysis. Craik (1973) explained "depth" in terms of the meaningfulness extracted from the stimulus rather than the number of analysis performed or the processing time.

Hyde and Jenkins (1973) performed a study with several orienting tasks to vary depth of processing in a incidental learning situation. One group performed the orienting tasks with no instructions to learn the material (incidental learning) another group had the orienting task with instructions to learn the words (intentional learning). There was also a control group that did not do the orienting task but received only intentional learning instructions. All of the groups were given a test of free recall shortly after the completion of the orienting task. The results indicated that retention after semantic tasks was 83% higher than the non-semantic tasks on a list of associatively

related words. They also found that incidental learners performing a semantic orientating task recalled as many words as intentional learners in the control group. This implied the nature of the processing activity determines how much is remembered rather than the intent to learn. Finally they found that there was no statistically significant difference between intentional and incidental learners who performed the same orienting task.

Motivational states are affected by a variety of internal and external factors. Intrinsic motivation is involved when the task activity itself provides the only source of reward, possibly in the form of enhanced feelings of competence and self-determination. On the other hand, extrinsic motivation is controlled by someone other than the individual themselves. These are incentives or rewards that provide satisfaction that is independent of the activity itself (Deci, 1975). Recently there has been increased interest in external factors relevant to motivation. Research into the effects of motivation on cognitive performance by varying the amount or reward offered for successful performance.

Feldman (1964) found that subjects offered high-incentives worked faster than low-incentive subjects on a digit-cancellation task, but made considerably more errors. This suggests incentive increases speed at the cost of accuracy, or quantity at the cost of quality. Eysenck (1984) states that incentive has various effects on cognitive performance: it increases performance speed, decreases accuracy of performance, produces some cognitive inflexibility, and increases short-term storage capability.

This study examines the effects of levels-of-processing and incentives on recall in an incidental learning situation.

Method

Four groups of twenty subjects were used. Each consisted primarily of fourthclassmen at the United States Air Force Academy ranging from between 19-21 years of age.

A portable audio cassette player, a tape with forty pre-recorded adjectives spoken at five second intervals, and response sheets with 40 spaces were used.

Group 1 Counted the number of vowels in words and was not offered an incentive to improve recall.

Response sheets were passed out to each subject with the following intructions:

"You will be presented a list of 40 words at 5 second intervals. Your task is to listen to the word and determine how many vowels are in the word and then write the numbers down on the sheet."

277

The cassette player was then turned on. After the tape finished, subjects were told to turn in their sheets and get a new one. Subjects were then instructed to recall all of the words they could in any order.

Group 2 Also counted vowels and was offered an incentive to improve recall.

Subjects in this group went through the same procedures as Group 1 with the exception of the instructions prior to the recall task. These subjects were instructed that they would receive "Coke privileges" for getting eighty percent correct on the recall task. This is a moderate incentive for fourthclassmen at the United Stated Air Force Academy. They were informed of this just prior to the recall test.

Group 3 Performed a personal relatedness task without recall incentive.

The sheets of paper were passed out to each subject with the following instructio.is:

"You will be presented a list of 40 words at 5 second intervals. I want you to listen to each of the words and determine how closely it relates to you, a 5 for relating a lot to you and a 1 for very little or no relation."

The same tape of 40 adjectives was then played. After the tape was played subjects were told to turn in their sheets and get new ones. Subjects were then instructed to recall all of the words on the tape that they could in any order.

Group 4 Also performed the personal relatedness task with a recall incentive.

Subjects in this group went through the same procedure as Group 3 with the exception of the instruction prior to the recall task. These subjects were also promised that they would receive "Coke privileges", for eighty percent correct on the recall test.

Results

The scores were compiled for each group, then each group's scores were averaged and compared. Means and standard deviation of number of words correctly recalled for each group were as follows:

Vowel w/incentive	$x = 11.0$, $s = 3.46$
Vowel w/o incentive	$x = 7.15$, $s = 3.40$
Relatedness w/incentive	$x = 19.80$, $s = 3.48$
Relatedness w/o incentive	$x = 15.55$, $s = 3.41$

In general, when the subjects were told to rate how much the words they heard related to them, they were able to recall more words than the subjects who were told to write down how many vowels each word contained. Offering an incentive also appeared to enhance recall.

278

The statistical results of the two-way ANOVA are as follows:

Table 1

Source	SS	df	MS	F	Fcv
Process	1487.81	1	1487.81	125.66	3.98
Incentive	314.12	1	314.12	26.53	3.98
Interaction	19.01	1	19.01	1.61	3.98
Within Cell	900.05	76	11.84		
Total	2720.99	79			

There is a main effect due to levels of processing and incentive, however the interaction is not statistically significant.

The number of incorrect responses for each group were also examined. Both levels of processing were compared to see if incentives increased the number of adjectives recalled that were not on the original list. Incentives caused a significant increase in the number of false responses for the groups that performed the vowel task (x_i= 4.75, x_c= 1.37, t_{38}= 3.93, p<.01). However, for the groups who had performed the relatedness task the increase in false alarms was not significant (x_1= 1.65, x_c= 1.00, t_{38}= .81, n.s.).

Discussion

The results supported the hypotheses. The level of processing brought about 95 percent higher recall than the vowel-counting task. The incentive had an additive effect of about half the magnitude. This indicates that learning can be enhanced with either an incentive or by relating the task to someone's interest. Also, the lack of interaction suggests each of these processes acts through different mechanisms. Incentive, however, may have detrimental effects on recall accuracy. Feldman (1964) explained that high incentive subjects worked faster on tasks but accuracy on performance decreased. Our result supports this view.

The positive effect of semantic processing is consistent with Craik and Lockharts (1972) levels of processing theory. Their assumption was that deeper processing created more elaborate, stronger traces. One problem identified by Eysenck (1978) is the vagueness with which depth is defined. Parkin (1979) claims that in many cases, the depth of processing of an orienting task is usually fairly obvious. In our experiment, rating how much an adjective relates to you would have been processed more deeply than counting the number of vowels in the word (especially in terms of the meaningfulness extracted from the stimulus).

Another factor to look at in our results was the incentive. The incentive while appropriate for the fourthclassmen would not have been sufficiently motivating for the few upperclassmen. However, the few upperclassmen probably experienced strong incentive to perform well in front of their subordinates. Something else to look at is the task itself. According to Revelle and Michaels (1976) tasks that have subjective probabilities of success between .1 and .5 are highly motivating. If the probability of success falls below .1 then motivation may be low no matter how high the incentive. Our specific task was free recall from a list of 40 words. Subjects were instructed that if they recalled 80% of the words they would receive the reward. Subjects may have been initially motivated but soon found that the goal would not be possible to attain so their motivation decreased.

In conclusion, our data supports the levels of processing theory and our hypothesis that incentive can increase recall. Our results showed two main effects: level of processing showed a 95 percent increase in recall with no increase in false responses, and incentives produced a 40 percent increase of recall, but introduced a two-fold increase in the number of false responses. There was no interaction between these two effects. If we assume that incentives affect conscious, intentional mental activity (Eysenck,1980) and the lack of interaction reflects independence then this study supports the position that the effect of processing level is not mediated by the intention to learn (Craik & Lockhart,1972).

References

Craik, F.I.M., and Lockhart, R.S. (1972) Levels of processing: A framework for memory research. Journal of Verbal Learning and Verbal Behavior. 11.

Craik, F.I.M., and Tulving, E. (1975) Depth of processing and the retention of words in episodic memory. Journal of Expermental Psychology: General. 104.

Eysenck, M.W. and Eysenck, M.C. (1980) Effects of processing depth, distinctiveness, and word frequency on retention. British Journal of Psychology. 71.

Eysenck, M.W. (1984) A Handbook of Cognitive Psychology. New Jersey: Lawrwnce Erlbaum Associates, Publishers.

Parkin, A.J. (1979) Specifying levels of processing. Quarterly Journal of Experimental Psychology. 31.

Revelle, W., and Michaels, E.J. (1976) The theory of achievement motivation revisited: The implications of intertial tendencies. Psychological Review. 83.

The Effects of Schema on Incidental Learning and Memory
ClC Mark C. Kuras
ClC Raymond M. Janssen
LTC David B. Porter
United States Air Force Academy

Abstract
This study used a video vignette to
examine the effects of schema and explicit
instruction on recognition, recall and
prediction accuracy. The results indicated
appropriate schemata promoted better
predictions only when subjects were given
advanced notice of the prediction task. There
were no other significant relationships
between schema appropriateness and indicants
of memory. Stopping the film clip and asking
subjects to make a prediction caused a
significant increase in the variance of
recall scores in comparison to the control
group.

This study looked at the relationships between
schema/scripts, verbal instructions, incidental learning,
and two indicants of memory: recognition and recall.
Schank and Abelson (1977) proposed that knowledge is stored
around many kinds of activities in knowledge structures
referred to as "scripts". Past studies have shown schema
(or scripts) can affect memory in interesting ways. One
effect is that scripts determine *how new material* is placed
in memory (Reed, 1982). Bower, Black and Turner (1979)
found that although appropriate scripts usually facilitate
comprehension and recall, they can also create errors. This
happens when subjects cannot distinguish between the
knowledge from the schema and the information presented
(Reed, 1982). Bransford and Johnson (1972) found that
schema facilitate the comprehension of complex passages and
events. Many memories are products of incidental learning.
Such learning is not due to intentional learning but is
incidental to other purposes (Glass & Holyoak, 1986).
Infact, Ross (1981) found that the more purpose subjects had
when faced with a task, the more detailed was their analysis
of inputs and the better these details were remembered.

We hypothesized that asking people to predict an
outcome from a scene in a movie about a city crime would be
more accurate, if the person was from a big city rather than
the country. People from the city should have schema better
suited for interpreting daily events in a city than people
raised in the country. Two specific hypotheses were tested.
We assumed more subjects from the city would accurately
predict an outcome in the film clip than the subjects from
the country. We also wondered if the relationship between
predictive accuracy and schematic propriety would be
affected by subjects' knowledge that they would have to make
a prediction. Our second hypothesis was that subjects who

were interrupted and asked to make a prediction about an outcome would show improved recall but not recognition. We suspected that invoking schema (which subjects should do in making a prediction) would improve recall by providing a framework but not necessarily memory for details (recognition).

Method

Subjects

Three groups of 15 cadet volunteers, between 17 to 21 years old, were selected. Both genders were used, but all groups were predominantly composed of males. Confidentiality of the results was assured by assigning subjects a discrete number code to be used on all forms.

Apparatus

The subjects were asked to rate themselves on a 7-point scale as being from the country (1) or city (7). Subjects viewed a five minute film clip from the television show Saturday Night Live. The television was a 26 inch Magnavox Color Digital Remote Control System. The video cassette recorder used to show the movie was a Fisher FVH 715. The prediction test was a single multiple choice question with six possible outcomes.

The recognition questionnaire contained fifteen multiple choice questions, each with four possible answers, relating chronologically to objective features in the film. The critical thing that defines a recognition task is that one is given an accurate representation of the information one needs to locate in memory (Glass & Holyoak, 1986).

In contrast, recall was tested by giving subjects instructions to list the main points of the movie. Unlike a recognition task, in a recall task, one is not given a representation of the information to be found in memory. Subjects must generate possible concepts (Glass & Holyoak, 1986).

Procedures

Three groups of 15 subjects each were used. Group 1 was the control group. This group entered the room knowing only that this was part of a psychological experiment. Then they were told they were going to watch a short clip from Saturday Night Live, but not told there would be a test following the movie. The video was then played to completion. Briefly, the clip was a story of the trials and tribulations of a jazz musician (Bradford Marsalis) in New York City. After arguing with his wife about finances, the musician takes his infant son and tenor saxaphone to a nearby street corner. After playing for several hours he gets into an argument with three young toughs who insist on playing rap music on the same street corner. He gets so angry, he chases them...only to have one of them circle back and steal his saxaphone case full of coins (this is the outcome other subjects were asked to predict). He returns to the corner, comforts his son, then continues playing collecting coins in his cap. His wife comes by and invites

him back home. At the end of the clip, subjects were given the tests for recognition and then recall.

Group 2 followed the same procedures used for group 1 with one exception. Near the end of video, the clip was stopped and subjects were asked to predict the next scene of the movie based on what they had already seen. Once this question was answered and collected, the clip was continued to completion.

Group 3 differed from Group 2 only in that this group was told before the start of the film that the clip would be stopped at some point and they would be asked to make a prediction. As with Groups 1 and 2, Group 3 did not know memory tests would follow the movie.

Results

T-tests were used to check the approximate equality of the three groups in distribution of city versus country self assessments. Unfortunately Groups 1 and 3 differed significantly ($t=3.65$, $df=28$, $p<.05$). Group 3 showed a greater number of city-oriented subjects than did Group 1. Group 2, however, did not differ significantly from either of the other two groups.

Although the proportion of subjects making the correct prediction in Groups 2 and 3 did not differ significantly ($t=.88$, $df=28$, $p>.05$), the relationship between subjects' schematic orientations and performance differed dramatically in the two groups. Point-Biserial Pearson Correlation Coefficients revealed that for the group that was surprised by the prediction task, r was $-.231$ ($df=14$, N.S.) but for the group of subjects who were told they would have to make a prediction, r was $.520$ ($df=14$, $p<.05$) which showed a moderate correlation between being from the city and obtaining the correct prediction. The difference in these two groups was also significant, $z = 1.976$, $p<.05$.

The means, standard deviations and multiple t-tests showed no significant difference between groups in their performance on the recognition test. The average number of questions correctly answered was 10.4 (of a possible 20) with a standard deviation of 1.42 questions.
When the recall data were tabulated, it appeared that there was a difference between the average correct responses in the three groups (control X=2.87, surprise X=3.23, informed X=3.73). Multiple t-tests were used to compare the three groups. The apparent differences were not statistically significant. However, the variance for Groups 2 and 3 was over twice the variance in the control group. $F(14,29)=2.45$, $p<.05$. This suggests that stopping the film and asking subjects to make a prediction (which was done for both Groups 2 and 3) significantly increased the between subject variation in recall performance.

Pearson correlations were used to determine if self-ratings (city versus country) had any effect on either recognition or recall. No significant correlations were found in any of the three groups. Recognition scores and

recall scores were also not correlated with each other (r=.09, n.s.). Point biserial correlations were used to compare prediction accuracy with recognition and recall scores. No statistically significant relationships were found.

Discussion

This study did not support our first hypothesis (i.e., that people from the city regardless of what group they were in would better predict than people from the country). Instead, the advantage of having an appropriate schemata only accrued to those who knew beforehand that they would have to make a prediction. Since Group 3 was given a purpose, city-oriented people appear to have employed their schema of city life from the beginning of the film. This enabled them to be better prepared for the prediction task. City-oriented people from Group 2 apparently did not use their schema and thus predicted poorly.

The second hypothesis, that making a prediction would invoke schema and in turn result in increased recall scores, was also not supported. There was no statistical difference between any of the groups on either recall or recognition scores. However, stopping the film and asking subjects to make predictions significantly increased the variance in recall scores. This suggests that overtly invoking schema had a positive effects on some subjects and a negative effect on others. This result is consistent with Bower's et. al. (1979) suggestion that the extra information, (ie., the input of prediction schemata) may actually increase error for some subjects. In other words, making subjects predict an outcome caused them to introduce their own information not related to the actual movie, and thus created a greater chance for error to occur on the subsequent recall task. The minimal correlation between recognition and recall scores suggests that our tests tapped relatively independent aspects of memory.

This research was based on Bartlett's (1932) classic study of memory for an unfamiliar story. We recognize that by checking so many differences and relationships, we have greatly increased our chances of uncovering "significant" results which are infact only spurious statistical artifacts. The general convergence of these findings with the existing literature, however, supports the validity of our experiment. Although the results we obtained were somewhat less than hoped for, the two findings which were significant seemed to be particularly interesting: subjects who are warned in advance of a prediction task make better use of the schema. Evoking schemata by asking subjects to make explicit predictions of ongoing events, significantly effects the way those events are subsequently recalled. This effect varies from one individual to another, but the overall effect is to increase variation between individuals.

References

Bartlett, F.C. (1932). <u>Remembering: A study in experimental</u> <u>and social psychology</u>. Cambridge: Cambridge University Press.

Bower, G.H., Black J.B., & Turner, T.J. (1979). Scripts in memory for text. <u>Cognitive Psychology, 11</u>, 177-220.

Bransford, J.D., & Johnson, M.K. (1972). Contextual prerequisites for understanding some investigations of comprehension and recall. <u>Journal of verbal learning</u> <u>and verbal behavior, 11</u>, 717-726.

Eysenck, M. W. (1986). <u>A handbook of cognitive psychology</u>. London: LEA.

Glass, A. L., & Holyoak, K. J. (1986). <u>Cognition</u> (2nd ed.). New York: Random House.

Klein, S. B. (1987). <u>Learning: Principles and</u> <u>applications</u>. New York: McGraw-Hill Book Company.

Reed, S. K. (1982). <u>Cognition: Theory and applications</u>. Monterey, California: Brooks/Cole Publishing Company.

Team Information Seeking

Elliot E. Entin, Ph.D.
ALPHATECH, Inc.
Burlington, MA 01803

Abstract

Teams of two physically separated decisionmakers engaged in information gathering on a set of closing unknown threats until, as a team, they felt they had enough information about the threats to accurately classify them as to type. The situation constantly required the decisionmakers to tradeoff the time spent seeking information to classify a threat with the accuracy to which that classification must be known. Results showed that three factors, team expertise, number of available probes for gathering information, and input load, had a significant impact on accuracy and timeliness of threat classification. The implication of these findings were discussed.

In the military domain, individuals on a command team are often not physically collocated. Those assuming the command and control obligations of a naval battle group are usually located on a number of distributed platforms. To exercise their command duties, these individuals must coordinate their activities over often limited and noisy communication channels. Situation assessment is an important task that all commanders share.. Using the resources and sensors under their control, commanders gather information and strive to formulate an accurate representation of the threat environment. Commanders can communicate with one another to increase their information base; share information, workload, and resources; solicit expertise; and coordinate actions. Such activities raise many interesting issues. Is it wiser for commanders to gather their own data or try to solicit information from others? How does the relative expertise of the commanders comprising the team impact this and other issue? For the teams, what variables influence the tradeoff between the time spent to gather information and the need to take actions? More specifically, how is it decided that enough information has been gathered and that the representation of the threat environment is accurate enough for effective processing of the threats.

There is much research literature on human situation assessment and information processing. However, a majority of this research has concentrated on a single decisionmaker or a group of decisionmakers in face-to-face contact. The present study investigates factors that influence information seeking for threat identification in distributed teams. The decisionmakers engage in information gathering on an unknown threat until, as a team, they feel they have enough information about the threat to accurately classify it as to type. The threats are moving and thus offer a limited processing window or time available before they penetrate a protected area and inflict damage. The decisionmakers are thus under pressure to acquire information about a target and decide its classification in a timely manner. In other words, the decisionmakers must balance the time spent seeking information to classify a threat with the accuracy to which that classification must be known.

Three factors affecting information seeking and threat identification were studied. The first dealt with the relative expertise of the members comprising the team; the second examined

the number of resources available to gather information; and the third studied the effect of input load (a form of workload) on the subjects' situation assessment activities. This was an empirical investigation and specific hypotheses based on theory were not made. However, from past experience it was expected that high team expertise, more resources to gather information, and lower input load would all foster better threat identification.

Method

Subjects - Eight graduate and undergraduate student volunteers from the University of Connecticut Engineering Department served as subjects. All subjects were paid for their participation.

Independent Variables - The experimental design manipulated three primary independent variables. Differential expertise varied the relative expertise of the two team decisionmakers over three levels. Differential expertise was created by varying the quality of information team members received concerning incoming threats. High experts received relatively noise-free high quality information, moderate experts received moderately noisy-moderate quality information, and low experts received relatively noisy-low quality information about threat attributes. A low expert would have to draw two samples from the threat's attribute distribution and combine the data normatively to match the quality of measurement afforded the moderate expert. Similarly, it would take two samples from the threat's attribute distribution for the moderate expert to match the information quality of the high expert. There were three threat types: "A," "B," and "C." For the H/L team condition, a high expert for "A" type threats but a low expert for "B" type threats was paired with a low expert for "A" type threats but a high expert for "B" type threats. The H/M team condition paired a high expert for "A" threats but a moderate expert for "B" threats with a moderate expert for "A" threats but a high expert for "B" threats. For the H/H team condition, a high expert for both "A" and "B" threats was paired with another high expert for both "A" and "B" threats. Throughout all the experimental conditions both team members were moderate experts for "C" threats. Number of probes available to gather information was the second independent variable and assumed two levels: one probe per decisionmaker and three probes per decisionmaker. The probe returned a two-dimensional measurement vector of the threat's attributes after a 30-second delay. Obviously, decisionmakers operating in the three-probe condition were able to collect more information in a given period of time than those operating in the one-probe condition. The third manipulated factor was input load. Three scenarios were designed drawing from a Poisson distribution to vary threat interarrival time within each scenario. Low input load was defined by a scenario that presented 16 threats and had an average threat interarrival time of 73.6 seconds. The scenario for medium input load presented 18 threats and had an average threat interarrival time of 63.5 seconds while the high input load scenario presented 21 threats and had an average threat interarrival time of 59.3 seconds.

Dependent Variable - Several dependent measures were assessed; the two best representatives of the teams' performance were the assessment score and the timeliness score. Assessment score was a combination of threat-type identification and the confidence subjects reported for that identification. If a correct threat-type identification was made, then a score of one was given and the assessment score was derived by multiplying the confidence rating by the task identification score. If, however, the threat-type identification was incorrect, then the one's complement of the confidence rating was determined and the resulting value divided by two (for the two remaining possible correct threats). Timeliness score was an inverse function of the elapsed time spent on threat processing. The measure started at one hundred percent when the threat appeared on the screen and declined linearly to a score of zero when the threat was no longer available for processing.

Experimental Environment - The experiment was performed with teams of two subjects. Each subject had a command station made up of a graphics display and an alphanumeric display terminal. The graphics display showed the temporal positions of the threats and the amount of resources available to each subject. The alphanumeric display presented clock time, communication messages exchanged between team members, threat information, and an input command line. Threat information consisted of the time a threat appeared on the screen, threat identification number (used for entering commands), current estimate of the threat attribute values, attribute values returned by a probe, current threat status, and resources used during a probe. The simulation was hosted on a PDP 11/60 computer.

Procedure - The two subjects comprising a team were physically distributed and communicated by way of scripted messages supported by the simulator. The message scripts allowed for requests, responses, and transfer of information and action items. All threats were displayed as an unknown "X" type that could be either "A," "B," or "C" threats. Threats appeared at the outer ring of the screen and moved toward the center. Each threat remained on the screen for six minutes unless processed. A threat had two associated attributes that were drawn independently from a truncated ($\pm 2\sigma$) joint Gaussian distribution, conditioned on the threat type. This required subjects to use both attributes in determining theat type. The experiment was structured so that each team member saw different information about the same threats and each team member was unaware of the others' actions unless they were specifically communicated.

All subjects were told that, as teams, their goal was to attain the highest possible accuracy in identifying threat types and attributes in the shortest possible time. A trial's activity began when one or several unknown threats appeared on the subjects' displays. After deciding which threat to work on, subjects were faced with three decision options: 1) when and how much to probe a threat, 2) when and what to communicate to their partner, and 3) when to stop collecting information and process (attack) a threat. Because threats were constantly moving toward the center, they offered a finite opportunity window. Teams continually strove to balance their information seeking activity (which affects team timeliness) with their desire to accurately identify the threats for effective processing. When a team decided to attack a threat, an estimate of threat type was made, and a confidence rating given on this estimation. Teams were presented with 16 to 21 threats per trial, for a trial duration of 20 to 23 minutes. At the end of each trial, team performance was displayed.

Design - The three independent variables were completely crossed to produce a $3 \times 2 \times 3$ three-factor within subjects analysis of variance design. Each team was randomly assigned to one randomized ordering of the twelve treatment conditions. Data were analyzed with the MANOVA routine of the SPSS$_X$ statistical package.

Results and Discussion

The expertise main effect showed that the highest assessment scores occurred when both team members were high experts ($p < .05$). It also appeared that replacing a low expert with a moderate one had little effect on assessment score; the performance for these two conditions (H/L and H/M) was about equal. There appeared to be a threshold for team expertise. Unless both members of the team were high experts on both A and B threats, team performance did not increase above the level set by the H/L team condition. Having both members of the team highly capable of processing both A and B threats allowed each team member to work somewhat autonomously, thus reducing intrateam overhead. This reduction in overhead workload allowed better team performance.

There were no main effects due to number of probes. However, the effect of input load on timeliness and assessment score was significant ($p < .02$ and $p < .005$, respectively). Teams were the most timely and earned the highest assessment score when input load was lowest. This was not surprising as workload was directly related to input load. In the low input load condition, subjects had sufficient time to plan their moves and consider information obtained. Subjects also had fewer threats to attend to (per unit time), thus allowing a quicker reaction to each and yielding a favorable timeliness score.

The expertise × input load interaction was significant for assessment score. Figure 1 shows a similar pattern of results for the H/L and H/M expertise conditions over the input load conditions. The H/H teams, however, showed better assessment performance in both the high and low input load condition. Performance for all expertise conditions was about the same for the moderate input load condition. A simple effects decomposition of the interaction revealed that the H/H expertise teams performed significantly better than the H/L and H/M teams in the low input load condition ($p < .03$) and in the high input load condition ($p \leq .05$), but were not different in the moderate input load condition. Thus, it appeared that when both team members were high experts for A and B threats, the team could take greatest advantage of a low input load or were best able to overcome the adversities of a high input load. Where workload was low and ample time existed to gather information via probes and communications, the superior information processing afforded by high expertise on both tasks lead to better performance. If workload was high and information gathering by probing and communication was at a premium, the ability of the H/H team to act antonomously resulted in lower intrateam overhead and better performance than in the other team conditions.

To further explore the data, an expertise × probe MANOVA was carried out within each input load condition for the assessment and timeliness scores. Figure 2 shows the resulting interaction within the low input load condition for timeliness. Having only one probe available reduced the timeliness of the H/L teams as compared to the H/M and H/H teams. However, when three probes were available, the H/H team was clearly more timely than either the H/L or H/M teams. Under low input load, all teams had more time to probe and collect information about threats. However, confronted with restricted information gathering resources (i.e., the one probe condition) those teams able to gather moderate to high quality information per probe had an advantage over the team receiving lower quality information. Thus, H/M and H/H teams felt more secure sooner about the nature of the threat and made timelier responses. When probe resources were not as restricted (i.e, the three probe condition) H/L and H/M teams probably attempted to probe more and thus became less timely.
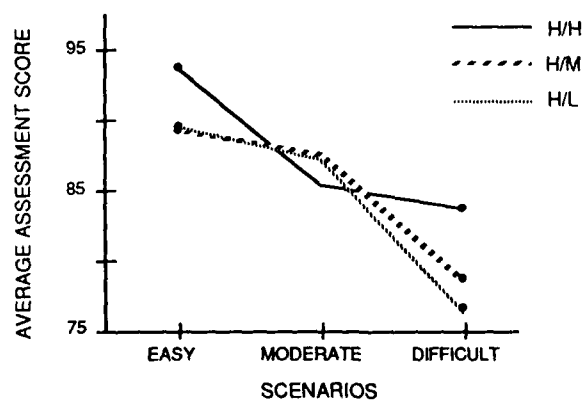
R-4641

Figure 1. Assessment Score as a Function of Team Expertise and Input Load.

289
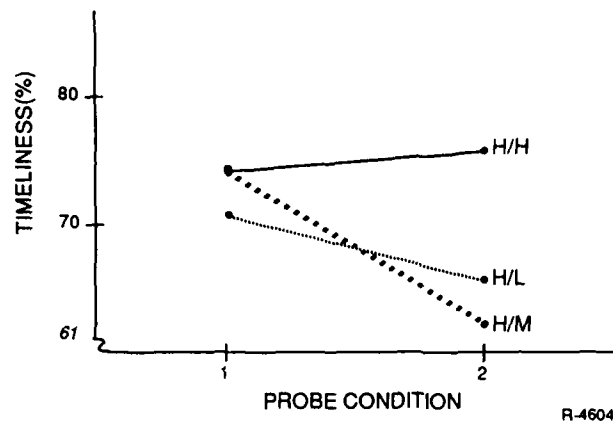
Figure 2. Timeliness as a Function of Expertise and Probe Condition for the Low Input Load.

Probabilistic Inference and Ambiguity in a
Missile Warning Officer Task[1]

Donald U. Robertson, Ph.D.
Tamara Della-Rodolfa, B.S.
Indiana University of Pennsylvania
John Forester, Ph.D.
AAMRL, Wright Patterson AFB

Abstract

The Einhorn-Hogarth model of probabilistic inference under ambiguity
was applied to probability judgments of attack in a missile warning officer
task. A revised model was also tested and provided better fit to the data.
Parameters from the revised model were then used to predict judgments in a
probabilistic information integration task. Results suggested that
judgments in a multiple cue probability task can be predicted from
ambiguity model parameters estimated from single cue judgments.

This paper described the results of applying the Einhorn-Hogarth ambiguity
model (Einhorn & Hogarth, 1985) to inferences about the probability of attack
in a missile warning officer task. In one version of the model, subjects anchor
on experimenter provided probabilistic information and adjust based on two
factors: the amount of ambiguity associated with the information and attitude
toward ambiguity under the circumstances. The model can be written as:

$$S = p + \Theta(1 - p - p^{\beta}),$$

where S is the predicted judgment, p a probability, Θ a parameter reflecting
ambiguity, and β a parameter based on attitude toward ambiguity. The value of
β also determines the crossover point, a value of p for which there is no
adjustment. A different formulation of the model, here called the revised
Einhorn-Hogarth (REH) model, is based on a linear rather than nonlinear
adjustment term and direct estimation of the crossover point. This model is
given by:

$$S = p + \Theta(pc - p),$$

where pc is the crossover point. Initial comparisons suggest that the REH
model may be superior to the Einhorn-Hogarth (EH) model under conditions of high
ambiguity (Robertson, Sussmann, & Della-Rodolfa, 1988).

Both models were applied to a missile warning officer task (Robertson,
Della-Rodolfa, & Forester, 1987). Subjects were given information about a
missile launch from the fictitious country THEM. Launch information was
displayed by "fans" of varying sizes which overlapped to a greater or lesser
degree with the country US. Fan size was presented as an index of sensor
system reliability (ambiguity) with larger fans representing less reliable
information. Fan overlap with US represented the probabilistic information,
and attitude toward ambiguity was manipulated by readiness state.

Initial tests of the models were successful and the present study was conducted to extend the ambiguity models to situations in which subjects were required to judge attack probability based upon two sources of probabilistic information. Subjects were given information about a launch which could have been an unmaneuverable Type 1 missile or a maneuverable Type 2 missile. Differential maneuverability thus made the projected impact region fans wider or more ambiguous for the Type 2 missiles.

Judgment of the probability that a particular launch was an attack on US, then, required integrating the probability of attack given that a Type 1 missile was launched with the probability of attack given that a Type 2 missile was launched. In order to test whether the ambiguity models could be used to describe the integration function, the information subjects were given about launch sites was varied. Four different sites were constructed based on the cross-classification of attack or test with Type 1 and Type 2 missiles. Site information varied the marginal probabilities of missile type (40% or 60%) and the conditional probability of attack or test given missile type (33%, 50%, or 67%). For each event in the task, subjects saw two fans. The fans were constructed around the same heading, but wider fans represented a Type 2 missile and narrower fans a Type 1 missile. Selecting different headings permitted us to construct stimuli so that the proportion of overlap with US was greater for Type 1 missiles half the time and greater for Type 2 missiles the other half.

The primary integration model tested was a version of an integration function called the theta-weight model which has been successfully used to predict how subjects combine probabilistic information from sources that differed in credibility (Robertson & Starr, 1988). The theta-weight model is similar to the equal probability model (Mellers, 1986), but uses ambiguity model parameters based on single cue judgments to derive scale values and weights for multiple cue judgments.

When applied to the missile warning officer task, the overlaps for the two fans, p_1 and p_2, are first transformed to subjective overlap values, S_1 and S_2, by either the EH or REH model. The subjective overlap values are then weighted by the amount of relative ambiguity to predict the double fan judgment, S. The full model using the REH parameters is given by:

$$S = [\Theta_2/(\Theta_1+\Theta_2)][p_1 + \Theta_1(pc_1 - p_1)] + [\Theta_1/(\Theta_1 + \Theta_2)][p_2 + \Theta_2(pc_2 + p_2)].$$

The values for Θ_i and pc_i are estimates obtained from single fan judgments. Fit of the theta-weight model was compared with fit from eight other models: (a) p_1, (b) S_1, (c) p_2, (d) S_2, (e) $b_1p_1 + b_2p_2$, where b_i are least squares regression weights, (d) $b_1S_1 + b_2S_2$, (e) $(p_1 + p_2)/2$, and (f) $r_1S_1 + r_2S_2$, where values of r_i are site information base rates. Relative fit of p_1 with S_1, p_2 with S_2, and $p_1 + p_2$ with $S_1 + S_2$ give an indication of whether transformation to subjective overlap improved prediction, while other comparisons reflect the adequacy of weights in the various models.

Method

Subjects and General Design

Twelve paid subjects individually completed a 90 minute session in which

they made 110 probability judgments. Twenty warm-up trials were followed by a block of 66 single fan trails from which ambiguity model parameters were estimated. The last 24 trails consisted of site information and two fans, one for Type 1 and one for Type 2 missiles, constructed around the same heading.

Stimulus Display

All stimuli were presented with an IBM AT microcomputer. The display consisted of three areas: (a) a 18.42 x 15.56 cm area located in the upper left of the display which showed two hypothetical countries, US and THEM; (b) a 5.40 x 18.42 cm area in the upper right of the display in which event information was shown; and (c) a 24.45 x 1.90 cm area at the bottom which was used to prompt the subject to enter his or her probability judgments.

Missile launches from THEM to US were represented by fans which varied in total size and amount of overlap with US and orginated from one of four sites. Fan size was indexed by the subtended angle. Angles of $20°$, $30°$, $40°$, and $50°$ were used with smaller angles reflecting better sensor system reliability.

A projected missile impact region was defined for each fan the the following way. Launch site fan angle and heading (trajectory) for a launch were selected. Next the impact region was defined by a minimum and maximum missile range. In terms of the 1000 x 1000 world coordinate system used to construct stimuli, fans had a minimum range of 300 and a maximum range of 600 measured from the launch site. Different proportions of overlap were generated by using different headings. A computer program was written to determine the actual proportion of the impact region that overlapped US.

Single fan launches were constructed by crossing the four fan sizes with 11 overlap percentages that ranged from 0% to 100% in 10% increments. However, presenting all 44 of these events at each level of attack base rate would have made the experimental session too long. Therefore, the 44 events at each attack base rate were divided into two sets with half the subjects making judgments for each. The 22 events used all levels of overlap and all fan sizes, with each subject making two judgments at each overlap percentage. For a specific percentage of overlap and attack base rate (33%, 50%, and 67%), half the subjects judged fans of $20°$ and $40°$, the other $30°$ and $50°$ fans.

Double fan launches consisted of two fans, one representing a Type 1 missile and the other a Type 2 missile. There were six trials at each of the four launch sites. On three trials, the percentage of overlap for Type 1 missiles was greater than the percentage overlap for Type 2 missiles, and the opposite for the other half of the trials. Stimuli were selected so that the difference in percentage overlap was approximately 10%. Three pairs of fan sizes were used: (a) $20°$ and $50°$, (b) $20°$ and $40°$, and (c) $30°$ and $50°$. Crossing the two patterns of percentage overlap for Type 1 and Type 2 missiles with the three pairs of fan sizes gave the six trials for each site.

Procedure

Subjects, who were seated in front of the IBM AT computer, were told that they would be given information about missile launches from a fictitious country, THEM, and would be asked to estimate the probability that a launch was an attack against the country US. Subjects were then given a copy of the

instructions to follow as the experimenter read them aloud. Launch information was coded on a report log as a means of insuring that subjects were attending to the relevant information. All probability judgments were entered on the keyboard.

After the 20 practice trials, the experimenter read the instructions for the 66 single fan trials. Subjects were instructed to continue completing the log but to make their judgments based not only the system reliability and percentage overlap information, but also on the attack base rate. Attack base rate was explained as being the probability that if a missile was launced from that site, it woud be on an attack trajectory. After the 66 single fan trails, instructions for the double fan events were read. Subjects were told to complete the log for both missile types for each event. Stimuli were displayed so that subjects were able to switch from one fan to the other as many times as they wished before making their probability judgment.

Results

The analysis consisted of two steps. First, judgments made in the single fan scenarios were analyzed in order to evaluate model fit and obtain the parameter estimates to be used to estimate subjects' judgments for the double fan scenarios. Second, the alternative models for the double fan scenarios were fit and compared. Each of these analyses were done for aggregate and individual data, but only the aggregate data are presented here.

Single Fans

Because all subjects did not make judgments for each fan size at each level of overlap, data for $20°$ and $30°$ (high reliability) fans were collapsed as were data for $40°$ and $50°$ (low reliability) fans. The average judgments across subjects were computed for low and high reliability fans at the three levels of attack probability. Thus, there were six aggregate data sets that were fit with the EH and REH models. Both models were fit with a SAS nonlinear, maximum likelihood regression program using a squared loss function.

Examination of parameter estimates and residual sums of squares for the EH and REH models revealed several patterns. First, as attack base rate increased the crossover point increased. Second, signal reliability had its largest impact on the crossover point, with higher crossover points for high reliability fans in all six comparisons (three comparisons for each model, one for each attack base rate). Finally, there appeared to be some further effect of reliability on estimates of Θ; the values of Θ were higher for low reliability fans in five of the six comparison. However, because each model was estimated from a small number of data points (11 in each case), there was considerable overlap in the 95% confidence intervals for the parameter estimates.

The residual sums of squares for the EH and REH model fits were compared and the REH model provided a better fit in four of the six comparisons. Furthermore, estimation of β in the EH model was particularly problematic. The asymptotic standard errors were so large that none of the values for β were significantly differnet from zero. In contrast, all values for both parameters in the REH model were significantly different from zero. Thus, the REH model appears to be preferable, both from the standpoint of the fit of the model to

the aggregate data as well as in terms of the precision with which the parameters can be estimated. As a consequence, the integration models employed the REH model ambiguity parameters.

Double Fans

For these analyses, the data were the average judgments of subjects for the 24 double fan events. Six events were presented from each of the four missile launch sites. Each of the nine integration models described earlier was tested with multiple regression procedures which permitted inclusion of a constant. Adjusted squared multiple correlations were used as one index of model fit and all regressions were significant (R^2 ranged from .875 to .947). Of greater interest is the pattern of correlations. Comparisons of models using percentage overlap with similar models using REH predicted values shows that transformation of the percentage of overlap with the REH model increased prediction. That is, the correlation for S_1 was higher than that for p_1, S_2 higher than p_2, and $S_1 + S_2$ higher than $p_1 + p_2$. Furthermore, the model that provided the best fit was the theta-weight model. The theta-weight model was further compared with the $p_1 + p_2$ model. A squared residual was computed for each data point using the $p_1 + p_2$ predicted values and tested against the theta-weight predictions with a paired t-test. The theta-weight model had smaller residuals ($\underline{t}(23) = 2.30$, $\underline{p} < .06$).

Discussion

The results suggest that probabilistic judgments can be modeled accurately by taking into account ambiguity. Formal two parameter ambiguity models are also useful in predicting how subjects combine probabilistic information. The major advantage of these ambiguity based models is that weights and scale values can be estimated independently of judgments made in the integration task itself.

References

Ei...orn. H. J., & Hogarth, R. M. (1985). Ambiguity and uncertainty in probaoilistic ingerence. Psychological Review, 92, 433-461.

Mellers, B. (1986). Test of a distributional theory of intuitive numerical prediction. Orangizational Behavior and Human Decision Processes, 38, 279-294.

Robertson, D. U., Della-Rodolfa, T., & Forester, J. (1987). Ambiguity and probabilistic inference in a missile warning officer task: Report No. 1. Unpublished manuscript, Indiana University of Pennsylvania.

Robertson, D. U. , & Starr, J. (1988, April). Extension of the Einhorn-Hogarth ambiguity model to multiple sources. Paper presented at the annual meeting of the Midwestern Psychological Association, Chicago, IL.

Robertson, D. U., Sussmann, M., & Della-Rodolfa, T. (1988, April). Revision of the Einhorn-Hogarth ambiguity model. Paper presented at the annual meeting of the Eastern Psychological Association, Buffalo, NY.

Modeling Expertise in Air Combat Maneuvering

Timothy E. Goldsmith, Ph.D.
University of New Mexico
Roger W. Schvaneveldt, Ph.D.
New Mexico State University

Abstract

This paper describes an effort to develop an artificial intelligence model (AI) model of pilot decision making in air-to-air combat maneuvering. The model, Air Combat Expert Simulation (ACES), uses a production system architecture to represent the selection of an air combat maneuver given the scenario of a 1-versus-1 engagement. A comparison of selections made by ACES with those of expert pilots showed that ACES agreed with the pilots as well as the pilots matched their own selections. Criticisms of ACES are offered and future directions for research discussed.

ACES is a computer simulation of expert fighter pilots' cognitive skill in air combat maneuvering (ACM). The model is intended to serve as an expert system for certain limited components of ACM. Additionally, the model is based on and tested by expert pilots and is assumed to have psychological relevance. Currently, ACES is directed at selecting single air combat maneuvers for particular airspace situations. The primary effort is aimed at developing an expert system model for selecting these maneuvers, realistically updating a simulated airspace with flight equations, and evaluating the potential of ACES as a training tool. This work has been reported in Goldsmith and Schvaneveldt (1987).

The ACES project has its roots in earlier work by Schvaneveldt and Goldsmith and their colleagues into the nature of cognitive skills of fighter pilots (Schvaneveldt, Goldsmith, Durso, Maxwell, Acosta, & Tucker, 1982; Schvaneveldt, Durso, Goldsmith, Breen, Cooke, Tucker, & DeMaio, 1985). This work aimed at discovering the underlying conceptual framework used by pilots in performing ACM. Both experts and novices were studied, and an attempt was made to uncover the cognitive dimensions on which these two groups differed. Conceptual structures were represented by multidimensional spaces and semantic networks. ACES built upon this earlier research by representing in a computer model the decision-making processes involved in selecting maneuvers in air combat. Techniques of knowledge engineering in AI and expert-novice research in cognitive psychology were employed to derive the model.

The ACES effort is designed to achieve several goals. A primary objective is to understand better the nature of the cognitive skills underlying pilot performance in ACM. Of particular interest are fighter pilots' planning and decision-making abilities, how these skills are acquired, and what distinguishes an expert fighter pilot from a less-skilled performer.

A second goal is to develop a computational formalism for representing and employing expertise in ACM. In this sense, ACES is an attempt to create an expert system. However, ACES is not aimed at

replacing the human expert. ACES, for instance, does not consider the perceptual-motor skills that are so critical in real-time performance of ACM. Instead, its sole focus is at a cognitive level. Also, an important aim of ACES is to consider the psychological validity of certain aspects of its computational formalism. In contrast, expert systems are motivated primarily by performance criteria and are not intended to be psychological models.

A third goal of ACES is to teach student pilots the cognitive skills required in ACM. ACES is expected to serve as a desktop training system to supplement students' academic training. It should be emphasized that ACES is not intended to teach the perceptual-motor skill required to actually maneuver aircraft. Instead, its teaching potential lies in its ability to convey to students the mapping of particular conditions onto appropriate actions.

Overview of ACES

Represented within ACES is an ACM situation consisting of two opposing aircraft (AT-38s) of equal capabilities. Each situation is a snapshot in time describing the positions, orientations, and airspeeds of the aircraft. ACES begins by assigning values to each aircraft, creating an initial situation. Next, each aircraft cycles through a series of stages that include selection, action, and updating. During selection, an action is chosen that is deemed appropriate for the aircraft to perform at that time. The major focus of ACES is on selecting appropriate actions. Action selection is guided by a model of the planning and decision-making skills of an expert pilot. After an action has been selected, the aircraft performs the chosen action. ACES then updates the situation to reflect the new aircraft values. ACES has each aircraft take turns cycling through these three stages.

ACES contains three different functional components: maneuver selection, maneuver execution, and a user interface. These components occur as distinct modules within the computer program. Maneuver selection is represented by a production system written in the programming language Prolog. This component selects a maneuver for a particular aircraft in a particular airspace situation. An airspace situation is defined to be a discrete snapshot in time of the aircraft known by the model to be engaged in combat. Although each situation is a discrete segment, a series of situations portrays the continuous nature of the engagement.

The maneuvers selected are always classical fighter maneuvers such as low yo yo and barrel roll. Each maneuver is represented by one or more production rules. The condition side of a production rule specifies conditions in the airspace appropriate for the selection of a particular maneuver. The action side of the rule selects the maneuver for execution. The database of the production system is a description of the current situation and contains airspace information such as aspect angle, angle off, indicated airspeed, and altitude.

The second component executes classical maneuvers known to ACES. An aircraft is flown through a maneuver by specifying for that aircraft a sequence of inputs to a set of flight equations. The inputs are specific values for a set of control variables. The model has two sets of control variables. One set consists of bank, g, and thrust; the other is made up of flight path angle, heading, and thrust. The use of one set, rather than the

other, is based on computational convenience. Occasionally, it is easier with one set of variables to indicate changes required by a maneuver in aircraft position and orientation. However, each set provides sufficient direction to the flight equations to fly an aircraft.

Each maneuver is decomposed into a series of one-second segments. Specific control variable values are then defined for each segment. These values are passed to the set of flight equations which actually fly the aircraft through the maneuver. Airspace information is updated after each call to the flight equations. Aircraft representation and maneuver execution is written in Pascal.

The third major component of ACES is responsible for displaying airspace information to and accepting input from a user. An airspace situation is portrayed by both text and graphics displays. Aircraft information such as heading, airspeed, and altitude and relative information, such as range and closure rate are provided in a table of text information. A three-dimensional graphics display depicts the current orientation of each aircraft. Viewpoints in the airspace may be changed by the user allowing various perspective views of the aircraft.

A user interacts with ACES through a hierarchical menu-based interface. After selecting an action from one menu level, a new menu with more specific options is made available. Eventually the user arrives at the point of entering the desired value. Some options provided to the user include initializing airspace conditions, altering the graphics display, checking weapons parameters, and selecting and executing maneuvers. A more detailed account of the user interface is provided in Goldsmith and Schvaneveldt (1987).

ACES currently runs on an IBM personal computer and its compatibles. Specific requirements of the system include a color-graphics video card and graphics monitor, at least 384K bytes of memory, and a math coprocessor.

Evaluating the Performance of ACES

The performance of ACES was tested by having both the model and expert fighter pilots select maneuvers for a common set of ACM situations. The experts were seven Air National Guard Pilots (GPs) stationed at Kirtland AFB, New Mexico. All of the GPs were experienced in air combat maneuvering.

A set of air combat situations for two opposing AT-38s were developed. A wide range of tactical positions for the airplanes were used. An IBM personal computer was used to convey task instructions, display situations, and collect subject responses. The GPs were tested individually at Kirtland AFB.

For each situation, the GPs were asked to select the most appropriate maneuver for their aircraft. After responding, the pilots were asked to select the second best maneuver if another maneuver was appropriate. If a pilot selected a second maneuver, he was asked to select a third best maneuver if another maneuver was appropriate. At any point after selecting the first maneuver, the pilot could indicate that no other maneuver was appropriate.

Each of the situations examined by the pilots was also evaluated by ACES. Maneuvers that had their conditions satisfied for a particular situation were placed in a conflict set. A conflict resolution strategy rank-ordered the members of the set. The three highest ranked maneuvers

were then selected and compared to the pilots' selections.

The basic question was how well ACES's selections agreed with the GPs' selections. The analysis compared the maneuvers selected by ACES with the maneuvers selected by the GPs for a given situation. Similarly, the performance of each GP was analyzed by comparing his selections to the selections of the other GPs. Two general types of analyses were performed. The first examined general agreement on which maneuvers were appropriate for each situation. A second type of analysis considered the order in which the maneuvers were selected. Here the interest focused on how similarly the selected maneuvers were weighted in importance.

The complete analyses are complex and are reported fully in Goldsmith and Schvaneveldt (1987). It was found, however, that the probability of obtaining, by chance, the number of matches in maneuver selections between ACES and the GPs was statistically significant. In summary, the performance of ACES was quite good and generally approximated the performance of the GPs.

Criticisms of ACES

Perhaps the major criticism of ACES at present is its limited knowledge of maneuvering; currently it knows only about classical basic flying maneuvers (BFM). Classical maneuvers are used largely to get into weapons parameters against a non-maneuvering target. Because they form the foundation for maneuvering in more realistic engagements, they are an important component of ACM training. Also, in the course of learning the classical maneuvers, a pilot learns to perform meaningful sub-maneuver components.

However, during an actual engagement, a skilled pilot is likely to perform only part of a classical maneuver. Once his opponent begins to counteract, the classical maneuver will likely no longer achieve its desired goal forcing the pilot to initiate some other action. Often this other action is not a complete maneuver but rather some component that is part of a classical maneuver. Examples of these components include pull up and roll out. The skilled tactical pilot, then, is able to effectively apply BFM using segments of several maneuvers in smooth succession. By restricting knowledge only to classical maneuvers, ACES is limited to a somewhat artificial type of engagement.

A related criticism is a lack of planning by ACES. During maneuver selection, ACES focuses attention solely on the current airspace situation. It fails to consider the engagement's immediate history or to plan ahead for subsequent actions. However, planning a course of action is crucial for intelligent behavior, especially in a rapidly changing task environment such as ACM. Obviously, in realistic engagements some degree of planning is required. In fact, skilled pilots attest to the importance of predicting future occurrences and acting accordingly. That ACES fails to plan is a serious weakness.

ACES is subject to technical criticisms. These faults are due primarily to developing the model on an inexpensive personal computer. First, because of the computer's poor screen resolution (640 by 200 pixels), only low fidelity graphics are possible. Several types of information would ideally be displayed to show graphically a changing airspace situation. Included here are relative size of aircraft, distance between aircraft, orientation of aircraft, and details of the actual

airframe. However, high resolution is needed to adequately display all of this information. ACES compensates by emphasizing orientation of the aircraft and some detail of the airframe while sacrificing fidelity for the other cues. Although this approach usually provides an adequate view, it requires some getting use to and occasionally leads to misunderstanding.

Second, a relatively long delay (approximately 1.5 seconds) exists between displays of successive airspace situations. Several time intensive operations occur during this interval including executing the flight equations and computing screen points for the graphics display. Although real-time simulation is not required for ACES, there are occasions when a user would benefit from viewing an animated sequence of aircraft movement.

Future Directions

It has become clear to us that a model of air combat maneuvering needs knowledge on maneuvering actions other than classical maneuvers. We hope to incorporate into ACES action units both more specific and more general than the maneuver. The need for more specific actions is due in part to the gulf that exists between a description of a classical maneuver, say given by an instructor pilot or a training manual, and a definition suitably detailed for its execution via flight equations. Micro-maneuvers would be independent meaningful units of action that specify elementary changes in aircraft's orientation or energy level such as pull up, roll over, unload, and extend. Classical maneuvers could then be defined by sequences of micro-maneuvers.

We also see a need for introducing macro-maneuvers into the model. Macro-maneuvers are action units that specify general goals in an air-combat engagement. Macro-maneuvers would be achieved, at least in principle, through performing some sequence of lower-level actions. Some examples of macro-manuevers might include increase closure, reduce angles, and break off engagement. In order to represent these types of multiple action levels in a model of ACM we are considering employing a network architecture of the parallel distributed processing (PDP) type (Rumelhart, McClelland & the PDP Research Group, 1986).

References

Goldsmith, T. E., Schvaneveldt, R. W. (1987). *Representing and Training Expertise in Air Combat Maneuvering.* Technical Paper No. AFHRL-TP-86-60, Williams, AFB, AZ: Operations Training Division, Air Force Human Resources Laboratory.

Rumelhart, D. E, McClelland, J. L., & the PDP Research Group (1986). *Parallel Distributed Processing: Explorations in the Microstructure of Cognition (Vol 1).* Cambridge, MA: MIT Press.

Schvaneveldt, R. W., Durso, F. T., Goldsmith, T. E., Breen, T. J., Cooke, N. M., Tucker, R. G., & DeMaio, J. C. (1985). Measuring the structure of expertise. *International Journal of Man-Machine Studies*, 23, 699-728.

Schvaneveldt, R. W., Goldsmith, T. E., Durso, F. T., Maxwell, K., Acosta, H. M., & Tucker, R. G. (1982). *Structures of Memory for Critical Flight Information.* AFHRL-TP-81-46, Flying Training Division, Air Force Human Resources Laboratory.

Use of Progressive Deepening in Battle Management

Klein, Gary A. & Thordsen, Marvin
Klein Associates, Inc.
Yellow Springs, Ohio   45387

Abstract

We have performed a series of studies during the last several
years, examining the decision strategies used by proficient decision
makers under naturalistic conditions. We have shown that there was
very little contrasting of options or systematic evaluation of options
on abstract dimensions. Instead, decision makers appear to rely on a
recognitional strategy we have termed a Recognition-Primed Decision
(RPD) model. The RPD model has been derived from research with
individual decision makers. This paper examines the generality of this
model to situations where there is a need to consider more than one
option at a time despite constraints such as time pressure, and to
situations involving distributed decision making. Evidence is
presented showing progressive deepening in operational planning
sessions conducted during Army C^2 exercises.

Recognitional Decision Making

In a series of studies (e.g., Klein, Calderwood, & Clinton-Cirocco,
1986; Calderwood, Crandall, & Klein, 1987; Taynor, Klein, & Thordsen, 1987)
we have shown the prevalence of recognitional decision making. We have
modeled the decision strategies we observed as Recognition-Primed Decisions
(RPDs). The experts we studied were using their experience to judge the
familiarity of problems, and to recognize the typical way of reacting to
those problems. The decision makers would evaluate the action that seemed
to be called for; the extent of this evaluation seemed to depend on the
amount of time available and the presence of atypical dynamics that needed
to be considered more carefully. In other words, when faced with a decision
point, our experienced subjects were able to recognize the situation,
recognize the typical reaction, perform some evaluation of the feasibility
of that reaction, and then carry it out. Only infrequently did they do any
deliberative comparisons of different options, looking for strengths and
weaknesses. They rarely used anything like a multi-attribute decision model
to reach their decisions.

The RPD model holds that the decision maker uses perceptual cues and
other types of knowledge to recognize the prototypicality of a situation,
thus clarifying the goals that can be accomplished, critical cues to
monitor, causal factors to track, the focal point of the situation
(primarily relevant to adversarial situations), and level of progress in
reaching objectives. In the simplest case, the automatic RPD, the decision
maker will recognize the situation and the most feasible reaction, and will
implement that action. Given enough time, the decision maker may also
choose to verify that the action will not run into any problems. In the
most complex case, decision makers will use a serial evaluation strategy,
progressively evaluating one option until an obstacle is reached. Only then
do they move on to a second alternative action, evaluating it until it

301

satisfices or until it, too, must be abandoned as unworkable. By recognizing the prototypicality, the decision maker will recognize an action queue of typical responses, headed by the most likely. These will be evaluated serially in descending order of typicality and either implemented, modified, or rejected. In addition, the recognition of typicality carries with it expectancies about how the situation will develop, and these will help the decision maker realize if the problem has been misidentified.

The strength of this RPD strategy is that it takes advantage of the expert's recognition of familiarity and ability to generate an effective option as the first one considered. The RPD strategy also means that the experts do not have to generate an exhaustive set of options and then have to evaluate them. In other words, the standard laboratory paradigm of behavioral decision making does not apply here. In the field, we did not encounter naive subjects performing well-controlled and context-free tasks. Instead, the people we studied were quite experienced, they were sensitive to a large variety of contextual nuances, the tasks were dynamic and showed frequent changes in goals, and the outcomes that eventuated from their decisions were of high impact. Poor decisions could have resulted in injury or death in many of the episodes we studied.

The RPD strategy has been described in greater detail in some recent reports (Klein et. al., 1986; Klein & Brezovic, in press; Klein & Brezovic, in preparation).

Option Evaluation

What happens when decision makers have to consider a set of options? The RPD model describes how proficient decision makers generate options, and how they evaluate single options, but not how they select from a set of options.

In order to formulate some hypotheses about option selection under conditions such as time pressure, we reviewed the chess protocols presented by de Groot (1978), who had studied chess grandmasters, masters, and experts. De Groot, who was himself a very strong player, had asked his chess players to think aloud as they attempted to solve very difficult chess problems. His work is a good contrast to ours, since chess players must generate a set of options and select the best one, whereas the skilled decision makers we have studied were typically not trying to generate more than one effective option.

We reviewed the 40 protocols reported by de Groot (1978). In each one, the chess player considered at least two options, and at most 11 options.

We found that they do not contrast the strengths and weaknesses of one move vs. another (a la Multi-Attribute Utility Theory). They do not use abstract evaluation dimensions such as "center control" to contrast moves. In the 40 protocols we studied, ranging from Grandmasters to mediocre players, we saw direct contrasts only a few times, fewer than s, and the contrast was about the overall value of the move not about its features.

Instead what happens is that they identify promising moves and work out what will happen if these are selected. This is the process of "progressive

deepening," a term coined by de Groot himself, from his description of these data.

As skilled chess players progressively deepen, they are trying to make a decision based on which move feels best, which move feels decisive. This is not as vague as it sounds. Here is what they mean by "decisive": a line of play that leads to a clear win of a piece, or, failing that, a line of play that leads to a positional advantage, or a line of play that leads to trouble. Anything else is seen as inconclusive. If they can find a good decisive outcome, they select the move.

They are also building up an emotional reaction. As they imagine the play getting deeper and deeper (to further depth of search), they notice aspects that make them enthusiastic or uncomfortable. They say things like "I would like to have this position in a tournament." Their emotional reaction is basically a summation of these specific concrete observations. They don't have to assign weights or become analytical. They just have to be sensitive to their emotional reactions. In other words, this is a context-bound evaluation in which the specific events of the game govern the evaluation rather than abstract dimensions such as center control or possibilities for a King-side attack. So, at the end of their probing, they have a set of moves they rejected because they were awful or indeterminate, and a set that are acceptable, and they pick the one that has engendered the greatest enthusiasm. And they do it without any direct comparisons of abstract qualities of moves. That is why a Multi-Attribute Utility Aid feels so foreign to users--it forces them to focus on abstract evaluation dimensions rather than on the specific strengths and weaknesses of the individual cases.

Now let us turn to tactical decision making to see if we obtain the same findings. Thordsen, Galushka, Young, and Klein (1987) studied Army battle planning sessions and obtained evidence for progressive deepening. The setting was a simulated Battalion-level command-and-control exercise conducted at Ft. Hood, Texas, using the Army Training Battle Simulation System (ARTBASS). The scenario included offensive and defensive operations in Germany. Data were collected during observations of a 5-hour planning session conducted by S-3 (operations). During the five hours, 27 decisions were identified and in only one case was there any comparison between options. The rest of the decisions, 26/27, conformed with a recognitional model of decision making.

The data were analyzed to study how the decision making occurred. We were particularly interested in the strategies used since this was an example of distributed decision making. The planning was done in small groups, headed by the officer in charge of operations, and in consultation with other staff members representing logistics, intelligence, Fire Support, personnel, higher level staff officers and so on. The data were plotted to show how goals were formulated and barriers identified. A clear finding emerged that the decision strategy showed the same type of progressive deepening that de Groot had reported for chess players. The initial identification of a situation included a recognition of typical reactions. The staff members helped to consider how the options would be carried out, and to identify pitfalls and barriers. The staff also suggested ways around barriers. If no way was found to overcome a barrier, the option was left in

an indeterminate state. Later on, the team might find a way to save it. Otherwise it was usually forgotten.

Conclusions

Traditional Army doctrine has emphasized the generation of sets of options and the systematic and comparative evaluation of these to select the best one. Previous data with tank platoon leaders (Brezovic, Klein, & Thordsen, 1987) has suggested that under field conditions decision makers do not usually follow this doctrine. Thordsen et al. (1987) provide even stronger evidence for recognitional as opposed to analytical decision making.

The use of progressive deepening strategies in many ways may be a more powerful decision strategy that systematic generation and evaluation of options. It is better suited to time pressure; it allows for the improvement of options while they are being evaluated; and it may take better advantage of the experience of proficient decision makers.

Many training programs emphasize systematic generation and evaluation of options. Many decision aids and decision support systems are designed to foster systematic generation and evaluation of options. It may be valuable to rethink the rationale behind such training and decision support system programs.

References

Brezovic, C. P., Klein, G. A., & Thordsen, M. (1987). Decision making in armored platoon command (KA-TR-858(b)-87-05F). Yellow Springs, OH: Klein Associates Inc. Prepared under contract MDA903-85-C-0327 for U.S. Army Research Institute, Alexandria, VA.

Calderwood, R., Crandall, B., & Klein, G. A. (1987). Expert and novice fire ground command decisions (KA-TR-858(D)-87-02F). Yellow Springs, OH: Klein Associates Inc. Prepared under contract MDA903-85-C-0327 for U.S. Army Research Institute, Alexandria, VA.

de Groot, A. D. (1978). Thought and choice in chess (2nd ed.). New York: Mouton.

Klein, G. A., & Brezovic, C. P. (In press). Recognition-primed decisions. In W. Rouse (Ed.), Advances in man-machine systems research. Vol. 5.

Klein, G. A., & Brezovic, C. P. (In press). Evaluation of expert systems. In S. J. Andriole and G. W. Hopple, (Eds.), Defense applications of artificial intelligence. MA: D. C. Heath & Co.

Klein, G. A., Calderwood, R., & Clinton-Cirocco, A. (1986). Rapid decision making on the fireground. *Proceedings of the 30th Annual Human Factors Society Conference*, 1, 576-580. Dayton, OH: The Human Factors Society.

Taynor, J., & Klein, G. A., & Thordsen, M. (1987). *Distributed decision making in wildland firefighting* (KA-TR-858(A)-87-04F). Yellow Springs, OH: Klein Associates Inc. Prepared under contract MDA903-85-C-0327 for U.S. Army Research Institute, Alexandria, VA.

Thordsen, M., Galushka, J., & Klein, G. A. (1987). *Distributed decision making in a command and control planning environment* (KA-TR-863(C)-87-08F). Yellow Springs, OH: Klein Associates Inc. Prepared under contract MDA903-86-C-0170 for the U.S. Army Research Institute, Alexandria, VA.

Human Detection of Visual Targets in Cluttered Backgrounds[1]

Theodore J. Doll, Ph.D.
Georgia Tech Research Institute
Georgia Institute of Technology

Abstract

The ability of human observers to detect simulated infrared targets in cluttered backgrounds was measured using a rating-scale detection task. The probability of detection (P_d) increased in a negatively accelerated manner with target luminance and varied inversely with range. Furthermore, the increase in P_d with target luminance was more abrupt at shorter ranges. There was also an interaction of range and background luminance variability. At long range, P_d decreased as the variability of the background luminance increased. However, at medium and short range, P_d was in some cases greater for more variable backgrounds. This result is explained in terms of the similarity of the spatial power spectral densities of the target and background. It is suggested that it may be possible to predict target detectability in complex, cluttered scenes based on the relative power of the target and background in bands the width of visual spatial frequency channels.

The process by which human observers recognize targets in complex, cluttered backgrounds is at present very poorly understood. Although quantitative methods exist for predicting the detectability of targets in homogeneous backgrounds, attempts to do so for cluttered backgrounds have had only very limited success (cf., Greening, 1976). A better understanding of this process could be extremely valuable, both in optimizing sensor systems to improve human detection performance, and in devising more sophisticated algorithms for automatic target recognition (ATR) systems. Such knowledge could also play a role in the development of more effective techniques of camouflage and concealment.

It is well known that luminance and size affect the detectability of targets (Ginsburg, 1986). An additional factor which has been shown to be important in basic research on visual masking is the extent to which the target "blends into" the pattern of the background. If the target and background pattern are made up of patterns that are similar in spatial frequency and orientation, then target visibility is greatly reduced. Specifically, gratings within ± 1 octave in spatial frequency and within ± 15 degrees in orientation interfere with one another; outside of these limits there is relatively little interference (Olzak & Thomas, 1986).

The present research examines the detectability of targets in cluttered backgrounds. The scenes used were simulated IR images, but the results apply to any sensor system which provides images for human interpretation on a flat, monochromatic display. In such systems, the viewing conditions and stimuli differ considerably from those usually employed in basic visual masking research. In particular, the observer scans the scene in a series of saccades rather than viewing the stimuli in a single fixation. In addition, the target

---

is a two-dimensional form rather than the one-dimensional sinusoidal gratings often employed in basic research. The objective of this study was to relate human detection performance in such imaging systems to physical parameters of the target and background.

Method

Each of four observers viewed a large number of simulated IR scenes, one-half with a target and one-half without. The target, when present, was square and uniform in luminance. The backgrounds were generated by modeling pixel intensity values as a random variable with Gaussian amplitude distribution and an exponential spatial autocorrelation function.

Experimental variables included: (1) the luminance of the target relative to the mean luminance of the background, (2) the variance of the background luminance, and (3) the range to the target (which determined the apparent size of the target and the correlation length of the background texture). The sides of the target square subtended a visual angle of 0.67, 0.33, or $0.17°$ for the 250, 500, and 1000 m ranges, respectively. The correlation length of the background was 0.50, 0.25, and $0.125°$ for the 250, 500, and 1000 m ranges, respectively.

Observers were individually tested in a quiet room, which was dark except for the light from the carousal slide projector used to present the test scenes. An IBM XT controlled the projector and recorded the data. Slides were projected onto a standard folding screen. The observer viewed the screen from a distance of 1.67 m. Test scenes subtended a visual angle of $17.2°$ degree in height and width.

Each test trial began with an audible warning tone to alert the observer. One second later, a test scene appeared on the display, and remained on for five seconds, or until the subject made a valid response, whichever occurred first. The observer scanned the scene briefly and indicated his or her confidence that the scene contained, or did not contain, a target by pressing one of four keys on the IBM XT keyboard.

Stimuli were generated on a Deanza 8500 monitor with 512 x 512 pixel resolution and 256 intensity levels. The 256 intensity levels of the display were mapped linearly into a temperature range of $40°$ C in the IR scene. The average temperature of the background was constant for all scenes at $20°$ C (127 intensity level units). The illuminance of the displayed backgrounds without targets varied only slightly, from 0.60 to 0.68 fc. The stimuli were photographed with a 35 mm camera using Ectachrome ASA 100 color slide film.

The set of test stimuli consisted of 768 scenes, divided into 12 blocks of 64 test scenes each. Scenes with and without targets were shown in a random sequence, and the target, if any, was located randomly at one of 12 locations in the scene. Each of the 384 different target scenes appeared once in this set; each of the 36 different non-target scenes appeared either 10 or 11 times in this set. The test scenes in a given block were homogeneous with respect to background luminance variability, but varied in range to the target and in target luminance. An additional four practice scenes with the same background variability were added at the beginning of each block. The 12 blocks were presented in a different random order in each 16 daily session for each subject. Each observer also completed a practice block of 68 trials at the beginning of the first test sesssion.

The observers were three male and one female students enrolled at the Georgia Institute of Technology, who were paid for their participation. Subjects ranged in age from 21 to 24 years and had corrected or uncorrected visual acuity of at least 20/25.

Results

Sensitivity (d's) for each of the 32 conditions were computed separately for each observer. The d's were then converted to probabilities of detection, P_d, in high-threshold theory. One need not accept high-threshold theory in order to use P_d. P_d is simply an alternative scale of detectability, and has the advantage of being more readily understood, especially in the applications and engineering communities. The conversion was done by taking as the high-threshold ROC that line which intersects the TSD equal-variance, gaussian ROC at the point where the response criterion, beta, is 1.0 (which is the negative diagonal of the ROC plot). P_d is the y intersept of the high-threshold ROC line.

This conversion procedure is equivalent to equating the areas under the high-threshold and TSD equal-variance, gaussian ROCs. Green & Swets (1966, p. 50) note that the area under the ROC curve (as a percentage of the entire ROC plot) is a distribution-free index of the detectability of the signal. P_d obtained by the conversion process is a monotonic function of the area under the ROC, and therefore it is also an index of signal detectability. P_d obtained by this conversion is the same value that would be obtained by computing P_d directly from the raw data (using the rational equation for the high-threshold model) if the observer were behaving so as to maximize the percentage of correct responses.

Figure 1 shows mean P_ds over the four observers. It is clear that P_d increases with target luminance, at least at the 500 and 1000 m ranges. The fact that P_d is high even for low target luminance at the 250 m range may be due to a ceiling effect. However, there is another possible explanation, which will be offered later.

For any given value of target luminance, P_d decreases with range. Thus there is a general rightward and downward shift of the curves with increasing range. It also appears that there is an interaction of range and target luminance. That is, P_d grows more rapidly with target luminance at the 500 m range than at the 1000 m range.

It was anticipated that it would be more difficult to detect targets as the standard deviation of the background luminance (S_b) increased. This expectation was confirmed at the 1000 m range. However, at the 500 m range, the inverse result occurred at two target luminances. That is, it was easier to detect targets in highly variable backgrounds than in more homogeneous backgrounds. In the middle plot of Figure 1, at a target temperature of 7° C above the background, P_d is slightly greater when $S_b = 10^\circ$ than when $S_b = 4^\circ$. At a target temperature of 4° C above the mean background, P_d is much greater when $S_b = 10^\circ$ than when $S_b = 2^\circ$ or 4°. Both of these reversals occurred for all four observers; hence it's unlikely that they are due to chance.

This result can be explained in terms of the power spectral densities (PSDs) of the target and background. As noted previously, the correlation length of the background decreases with range, which means that at long range there is comparatively greater energy at higher spatial frequencies. The target size also decreases with range, reflecting the increase in its fundamental spatial frequency. Note, however, that the difference between the background correlation length and the target size is smaller at 1000 m range than at 250 m range. This means that the PSDs of the target and background are more similar at long range than at short range.

This explanation is illustrated in Figure 2, which represents spatial variations in luminance along a "slice" through the target for four different
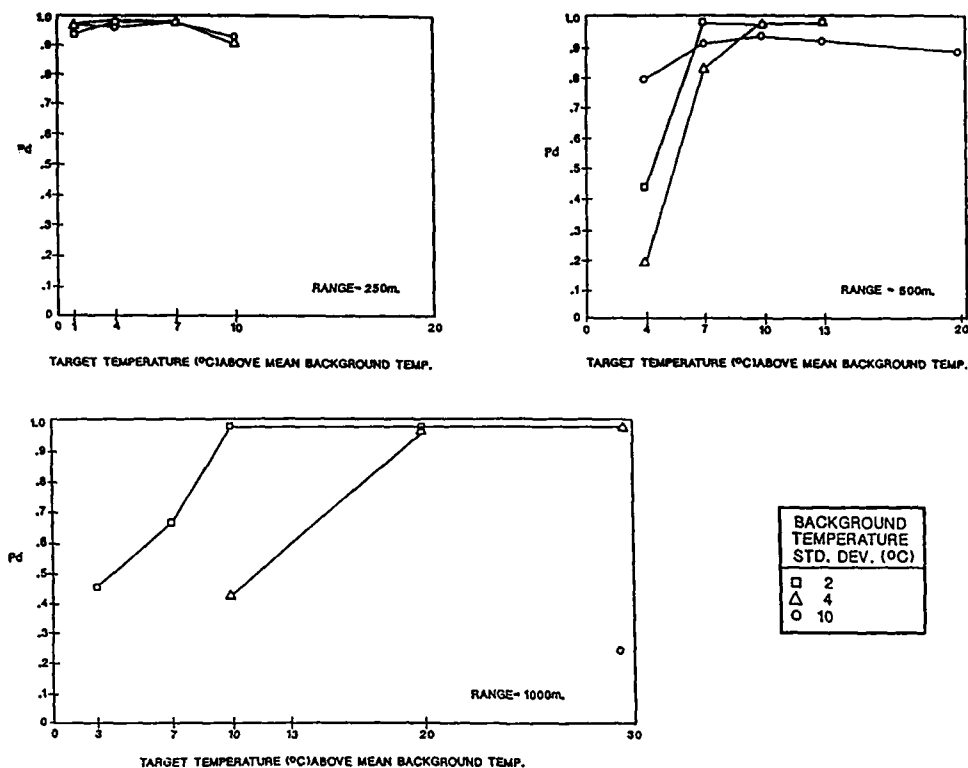
Figure 1. Mean probability of detection (P_d) as a function of simulated range, target temperature (luminance) and background variability.
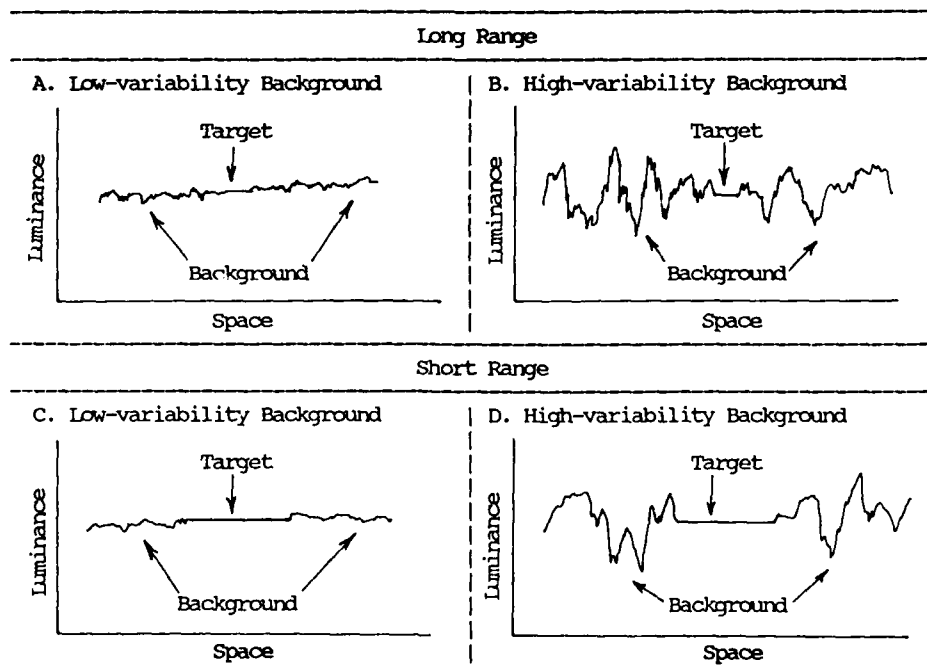
Figure 2. Spatial variation in luminance along a line passing through the target.

conditions. At long range (top-most frames), the target is similar in size to the spatial variations in background luminance. When the target luminance just above the average of the background, one can see why the target is easier to detect in a homogeneous background (frame a) than in a highly variable background (frame b). At short range the target is larger than the spatial variations in background luminance. In this case the target stands out more in the highly variable background (frame d) than in the homogenous background (frame c).

The disimilarity of the PSDs also explain why the target was always easy to detect at the 250 m range, even when its luminance was only 1° C above the average background luminance. At this range, the target size is even greater relative to spatial variations in background luminance than at the 500 m range.

Conclusions

The results show that the detectability of targets in cluttered backgrounds increases directly with target luminance, provided of course that detectability is not already at ceiling value due to other factors. The results indicate that an additional factor which influences detectability is the similarity of the PSDs of the target and background. Note that this conclusion refers to the relationship between the spatial frequency composition of the target and background, not to the spatial frequency of the target per se. The present study extends the findings of basic research on visual masking by showing that the relative spatial frequency composition of the target and background is also an important factor for more complex, dynamically viewed stimuli.

These findings suggest that the visual filter concept (cf., Olzak & Thomas, 1986) might be profitably applied to predict the detectability of targets in complex, cluttered, dynamically viewed scenes. It might be possible to develop an index, such as the maximum signal-to-noise ratio (SNR) in a specific bands, that predicts target detectability. The SNR-type index would be calculated from the PSDs of the target and background after they are multiplied by the visual contrast sensitivity function.

References

Green, D. M. & Swets, J. A. (1966). Signal detection theory and psychophysics. New York: John Wiley and Sons.

Greening, C.P. (1976). Mathematical modeling of air-to-ground target acquisition. Human Factors, 18, 111-148.

Ginsburg, A.P. (1986). Spatial filtering and visual form perception. In K.R. Boff, L. Kaufman, & J.P. Thomas (Eds.) Handbook of Perception and Human Performance. New York: John Wiley and Sons.

Olzak, L. A. & Thomas, J. P. Seeing spatial patterns. In K. R. Boff, L. Kaufman, & J.P. Thomas (Eds.) Handbook of Perception and Human Performance. New York: John Wiley and Sons.

Age and Vigilance: The Effects of Event Rate and Task Pacing

Jack D. Mohney, Captain, USAF
California State University, San Bernardino

Abstract

Age, background event rate (BER), and pacing effects were examined in this vigilance study. Volunteers ages 18 to 76 responded to infrequent critical stimuli presented on a lighted bar display. Analyses provided data that can be used as decision criteria for systems using multi-age populations.

American corporate and governmental agencies are faced with two significant, interacting phenomena. First, the U.S. population is growing older and is projected to age dramatically past the turn of the century. At the same time, the number of younger people is declining. These demographic trends not only affect our labor force composition, they also impact our present and future military resource pool. Second, our aging workforce must also cope with expanding machine complexity. As systems become more automated, the human becomes less interactive and more of a manager. Nevertheless, countless systems still fail when automated devices malfunction or some out-of-tolerance but rare condition happens. The Soviet Chernobyl nuclear accident is a vivid example of ignored critical signals and untimely or incorrect actions.

Past research has reported age differences in vigilance performance with older individuals consistently performing worse than their younger counterparts (Czaja and Drury, 1981; Davies and Davies, 1975). However, no study has examined the effects of event rate or pacing on these age differences (Parasuraman, 1984). BER also significantly affects vigilance performance (Parasuraman, 1979; Saito and Tanaka, 1977; Weiner, 1977). However, Weiner (1984) found that increasing task load during a monitoring/tracking task increased total vigilance performance. An optimal arousal level may therefore be present with a possible age function. It's also generally believed self-pacing is more efficient and less likely to show a performance decrement (Eskew and Riche, 1984; Drury, Morawski, and Tsao, 1979). Finally, Welford (1981) hypothesized the poorer performance of older people is caused by too much concentration on task response characteristics and not enough on the task itself. This compares to the widely-held notion that overly-rapid pacing is to blame.

This study investigated the influence of pacing, age, and BER on vigilance task performance. It was theorized that general performance declines with age, yoking, and time-on-task (TOT) with older subjects responding to self-pacing by decreasing baseline BER.

Method

Subjects

The subjects for this study were 36 volunteer men and 12 volunteer women ages 18 to 76, divided a priori into two age categories, 18-38 and 53-76. They were a mixture of government service, active duty military, civilian, and civilian retirees. Each had at least correctible 20/40 vision and no experience with a prolonged vigilance task situation. This 2 x 2 factorial design was counterbalanced with respect to age and sex. Subjects

were randomly assigned to one of 4 groups: younger with younger yoked (y-y), younger with older yoked (y-o), older with younger yoked (o-y), and older with older yoked (o-o). The independent variables were Ss' age, sex, time on task, and pacing (self versus yoked). The dependent variables were each S's reaction times, false alarms, misses, and background event rates.

Apparatus

The display consisted of a flashing bar of light with 3 white, 4-cm square translucent blocks located approximately 1/2 meter from the Ss' eyes, 30 degrees below horizontal line-of-sight, 23%, 47% and 55% additive contrast ratios respectively. Both Ss also had a telegraph key situated on each table. Room #1 contained an additional control box that allowed S #1 to manipulate the BER. By pressing and holding either the Faster or Slower button, S #1 could speed up or slow down the rate exactly 5% for each presentation. The display in Room #2 was yoked to the display in Room #1.

The apparatus operated through a Coulbourn logic network controlled by an Apple 2e computer. For each session, the computer began with a baseline BER of 60 signals per minute. Upper and lower limits of 130.972 and 46.427 signals per minute were built into the computer software with critical signals occurring randomly at a signal to noise (S/N) ratio of .008.

Procedure

Two paired subjects participated in one of 24 sessions, the only experimental deception being that neither subject knew of the yoked event rate. They were instructed to hold their index finger on a dot located 2.240 cm in front of the telegraph key and respond to the display by pressing the key when all three lights flashed on. Both subjects performed a one minute practice session prior to the one hour experimental session which was run a priori between 1000 and 1630 hours to control for the usual period of maximum alertness for humans.

Results

One subject (an older S yoked to an older self-paced S, o-o) voluntarily discontinued after approximately 2 minutes into the task. The mean of the median RT scores for that S's group, o-o, was substituted into the data to facilitate statistical analyses.

Figure 1 is four graphs of the data from the 24 trials. A three-factor Analysis of Variance (ANOVA) was performed on the RT data with pacing used as a repeated measure. Though younger Ss had quicker RTs than the older Ss, $t(46) = 9.496$, $p < .001$, the effect of pacing wasn't significant. Collapsed over time, younger self-paced Ss had quicker RTs than older self-paced Ss, $t(22) = 7.448$, $p < .001$ and younger yoked Ss had quicker RTs than older yoked Ss, $t(22) = 6.356$, $p < .001$. There was no significant difference in RTs between older self-paced Ss and older yoked to younger Ss and older yoked to older Ss. There was also no significant difference in RT differences of older self and older yoked subjects and RT differences of younger self and yoked Ss. While there was no significant difference in RT between older Ss yoked to either older or younger Ss, the younger Ss yoked to the older Ss had quicker RTs than younger Ss yoked to younger Ss, $t(10) = 2.641$, $p < .05$.

Linear trend analysis of TOT on RT indicates that while RT collapsed over groups between successive 10 minute time intervals wasn't significant, there was enough of a gradual, linear increase to cause significant differences between the initial, mid, and final time periods.

Due to low, erratic false alarm (FA) data, FA analysis involved only differences in means. Yoked Ss had more false alarms than self-paced Ss,

t(46) = 2.959, p < .010. The only FA significant difference occurred when a younger S was yoked to another younger S, t(10) = - 2.494, p<.050. Also because of low data, miss analysis involved only differences in means and linear trend analysis. Yoked Ss had more misses than self-paced Ss, t(46) = 5.106, p < .001.
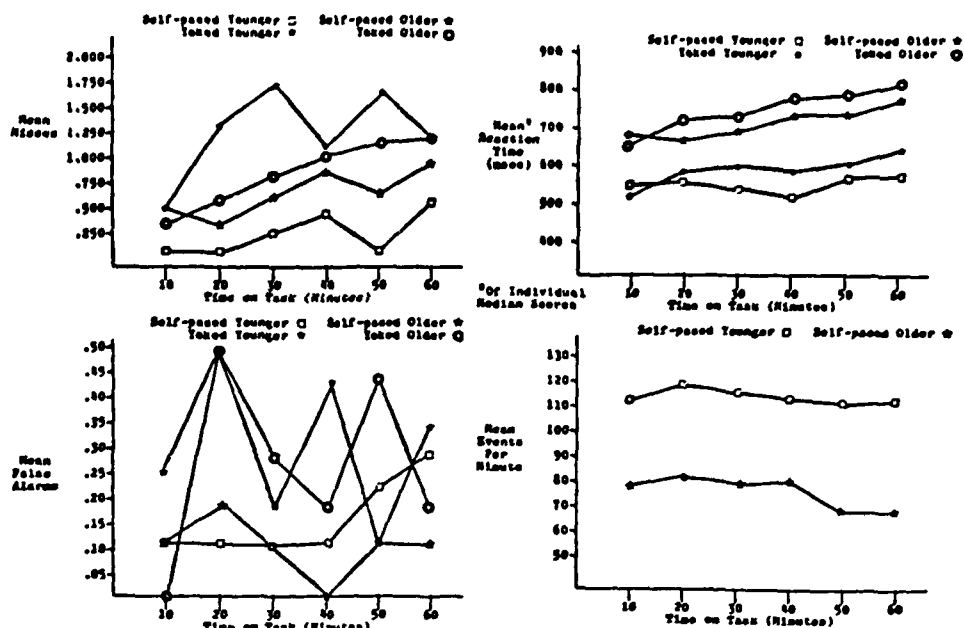
Figure 1.   Mean Misses, FAs, RTs, and BERs for TOT.

Yoking was a factor in all situations except when an older S was yoked to another older S. Trend analysis of TOT on misses indicates that while misses increased linearly with TOT and an age of self-paced by age of yoked by TOT interaction, there were no separate age by TOT interactions.

A separate 1 between, 1 within ANOVA was performed on BER data. Collapsed over time, younger self-paced Ss chose quicker BERs than older self-paced Ss, t(22) = 14.051, p <.001. There was no significant correlation between RT and BER (collapsed over all groups), r = .017, t(22) = .080.

Using signal detection analytical methods outlined by Kantowitz and Sorkin (1983), d' and the associated criterion values were calculated and summarized on the Receiver Operating Characteristic (ROC) graphs in Figure 2 for subject age, TOT, and pacing to emphasize d' and (β) differences. However, these ROC curves are only approximations based on a single data point because group differences were manipulated rather than the perceptual and response
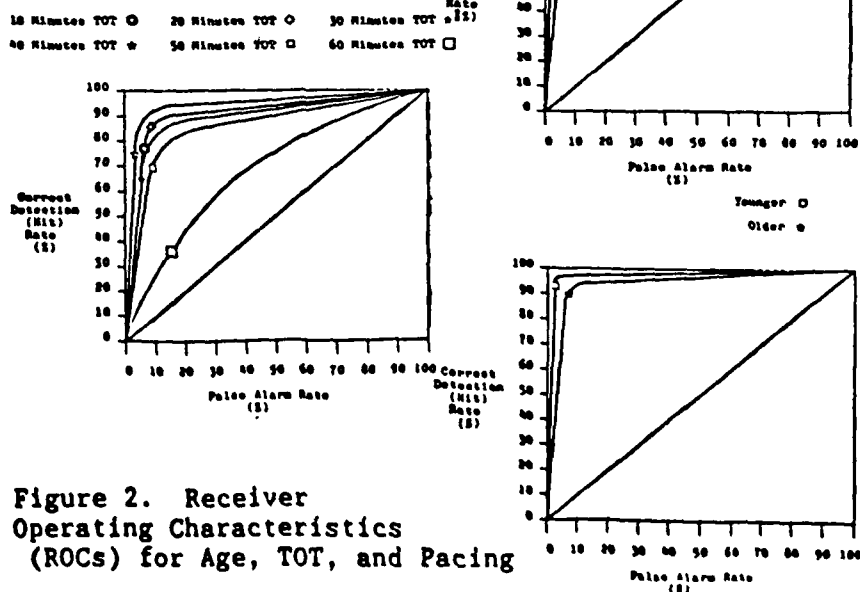
Figure 2.   Receiver Operating Characteristics (ROCs) for Age, TOT, and Pacing

criteria within each group. It was, in essence a snapshot of these variables.

Discussion

Overall, older Ss had longer RTs than their younger counterparts by 26.593%, lending support for baseline differences in d' and (β). Also, there was no significant difference in missed critical signals.

The effect of pacing on vigilance, while not total, was strong. It affected the number of misses and false alarms, but not RT. This is further reinforced by the 21.875% increase in the d' of self-paced Ss over yoked Ss.

As predicted, RT increased over time on task. Between 10 and 60 minutes TOT, RT increased 116.838%, FA by 200.000%, and misses by 276.836%, all statistically significant. Also, between 10 and 60 minutes TOT there was a 54.310% d' degradation.

Self-pacing didn't affect performance in all areas. RT wasn't affected and in fact, younger self-paced subjects had a better performance increase in miss and FAs than their older counterparts. There was also no age affect on TOT.

As postulated, older Ss did choose a slower BER than their younger counterparts, reinforcing the concept of a baseline d' difference between older and younger Ss. Decreasing the event rate compensates for this difference. Finally, there was no time effect on BER.

Since the level of vigilance varies directly with arousal changes induced internally through boredom or externally through outside interaction, variations in a person's decision criterion (β) affect their time-related vigilance performance. These criterion changes reflect changes in a person's expectancy of a critical signal. In this experiment, these changes were a function of age, pacing, and to some degree, BER.

For the HFE specialist or corporate/agency decision maker, this has important ramifications. If older people are to be effectively integrated into evolving monitoring task situations, their capabilities should influence system configuration. First, their slower reaction time must not cause catastrophic system failure. Second, self-pacing is preferable if detection and false alarm rates must be tightly controlled and minimized. If machine pacing is required, older people may integrate better into the system because of their smaller performance decline when yoked. Third, the event rate should be variable so older monitors can compensate for their initial d' decrement and the associated time-related decrement. Finally, the specific design of the displays and response modalities are most crucial to the performance of older monitors and should have high discriminability and ease of execution.

In summary, future vigilance scenarios with older monitors should have displays with high signal/noise ratios, variable event rates, involve self-pacing, and incorporate simple response modalities that aren't critically reliant on reaction time.

References

Czaja, S. & Drury, C. (1981). Aging and Pretraining in Industrial Inspection. Human Factors, 23, (4), 485-494.

Davies, A. & Davies, D. (1975). The Effects of Noise and Time of Day Upon Age Differences in Performance at Two Checking Tasks. Ergonomics, 18, 321-336.

Drury, C., Morawski, T. & Tsao, Y. (1979). Human Performance in Sampling Inspection. Human Factors, 21, (1), 99-105.

Eskew, R. & Riche, C. (1982). Pacing and Locus of Control on Quality Control Inspection. Human Factors, 24, (4), 411-415.

Kantowitz , B. & Sorkin, R. (1983). Human Factors: Understanding People-System Relationships, New York: John Wiley and Sons.

Parasuraman, R. (1979). Memory Load and Event Rate Control Sensitivity Decrements in Sustained Attention. Science, 205, 924-927.

Parasuraman, R. (1984). Sustained Attention in Detection and Discrimination. In R. Parasuraman and D.R. Davies (Eds.), Varieties of Attention. San Diego: Academic Press.

Saito, M. & Tanaka, T. (1977). Visual Bottle Inspection Performance in Highly Paced Belt-Conveyer Systems. Journal of Human Ergology, 6, 127-137.

Weiner, E. (1977). Stimulus Presentation Rate in Vigilance. Human Factors, 19, (3), 301-303.

Weiner, E. (1984). Vigilance and Task Load: In Search of the Inverted U. Human Factors, 26, (2), 215-222.

Welford, A. (1981). Signal, Noise, Performance, and Age. Human Factors, 23, (1), 97-109.

Situational Awareness: A Conceptual and Methodological Framework
Kelly Harwood, Barbara Barnett, and Christopher Wickens
Institute of Aviation, University of Illinois

Abstract

Situational awareness is a unique concept that refers to the pilot's knowledge of a dynamically changing situation. The need to understand the concept is growing ever important as technology proceeds towards highly complex automated systems like virtual work environments or cockpit-based expert systems. A major challenge is to develop paradigms in which situational awareness may be examined and to operationalize research questions that will meaningfully lead to system improvements. This paper suggests a conceptual and methodological framework for considering situational awareness.

The Representation of Situation Awareness

First, what is it? Stein (1986) defines situational awareness as "the knowledge available to the pilot on critical matters such as the overall tactical situation, his own mission profile, weapons status, the positions and objectives of friendly aircraft, disposition and apparent objectives of hostile flights, presence of threats, refueling rendezvous and other mission data" (p. 113). Stiffler (1987) uses the definition "the ability to envision the current and near term disposition of both friendly and enemy forces." Fracker's (personal communication, 1987) definition highlights the recognition of particular states of dynamic variables in a given scenario, and the formulation of a diagnosis regarding the processes underlying these present states as critical aspects of situational awareness.

Four dimensions distill out from these definitions: 'where', 'what', 'when', and 'who'. Where refers to spatial awareness, that is, the pilot's knowledge of his location in space and of the spatial relationships between objects. What characterizes identity awareness, or the pilot's knowledge of the presence of threats and their objectives. It also includes the pilot's awareness of system state variables such as engine status and flight performance parameters. Who is associated with responsibility or automation awareness; that is knowledge of "who's in charge". Finally, When signifies temporal awareness and addresses knowledge of the occurrence of events as the mission evolves. These four dimensions can be configured to form a 'pseudo' pyramid. 'Where', 'what', and 'who' lie on one plane and form the triangular base of the pyramid, while 'when' extends out along the z-axis.

The special status of the temporal component in this representation reflects the fact that time is the 'hallmark' of situational awareness. The past is important for disambiguating the present, and the past and present must be used to predict the future. The relationship of time with the other three components is what sets situational awareness apart as a unique concept. These relationships are depicted by links 1, 2, and 3 in Figure 1. Link 1 connects 'when' and 'who' and signifies the changing of responsibilities between machine and pilot. Link 2 connects 'when' and
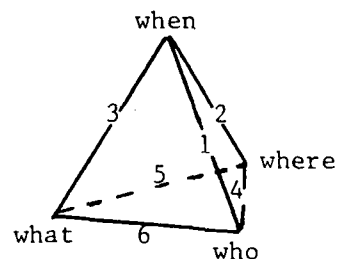
Fig.1. Situational Awareness

316

'where' indicating changes in spatial awareness during navigation. Link 3 connects 'when' and 'what' and refers to changes in knowledge of threats, friendly aircraft and system state variables as the mission unfolds. Temporal awareness is the common denominator among the other three dimensions; hence its position at the apex of the pyramid.

Given the unifying role of 'time' there are also interactions between the dimensions -- 'where', 'who', and 'what'. These interactions are indicated in Figure 1 by links 4, 5, and 6. Thus we can conceive of the relationship between 'who and where' (link 4) as entailing the activation of some automated system at a certain location in flight; for instance the activation of the flight director at the commencement of an instrument approach. The relationship between 'what and where' (link 5) might involve locking in on a refueling aircraft when over a specific geographic location. Finally, link 6 signifies the relation between 'what and who'. Such a relationship may be characterized by the activation/deactivation of an automated sighting system when in the vicinity of friend or foe. Tne interactions between and among the four dimensions are clearly complex and have far reaching implications for the design and configuration of cockpits, missions, and training programs.

Ultimately, situational awareness itself is a dynamically changing entity. Thus, over time, one would expect that certain 'links' in Figure 1 might be emphasized over others. (In other words the length of each link signifies the relative importance of the various relationships between dimensions.) These changes in importance over time can be depicted by having the pyramid take on different shapes. These shapes signify changes in situational awareness as the mission evolves. For two-way interactions, the pyramid decomposes into a straight line. For example, spatial awareness may be a priority at a certain point in the mission thus emphasizing the relationship between 'where' and 'when' (See Figure 2a). A slightly more complex situation would involve the relationship between three dimensions. For example, a rendez-vous with a refueling aircraft involves awareness of the relationship between 'where', 'when' and 'what'. This situation is depicted as a single 'face' of the pyramid as shown in Figure 2b. Figure 2c shows the
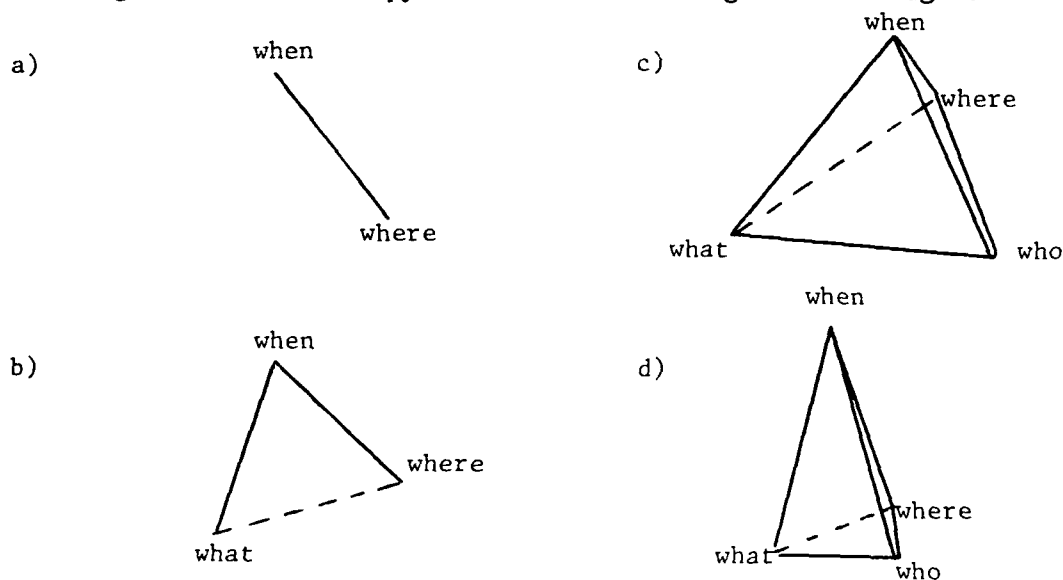
Figure 2. Relationships Characterizing Situational Awareness

situation of activating an automated weapons system to lock on to a target. Here the dimensions of control/responsibility (who) and identity awareness (what) are most important. Spatial awareness is of lesser importance assuming the pilot has the target in sight and is locked on to it; hence the shortened link between 'when' and 'where' (see Figure 2c). However, if the threat moves out of the pilot's sight, spatial awareness suddenly increases in importance and 'situational awareness' takes on a new 'shape' as shown in Figure 2d. Going from Figure 2c to Figure 2d also captures the dynamic sense of situational awareness.

Relevant Methodologies and Approaches to Capturing Situational Awareness

A variety of approaches is needed to fully explore the nature of situational awareness. One can envision these approaches as varying in the extent to which they tap explicit or implicit situational knowledge. At the 'explicit' end of the measurement continuum is the 'probe technique' for tapping situational knowledge at periodic intervals during flight (Marshak, Kuperman, Ramsey, & Wilson, 1987). In the middle, is the embedding of specific tasks for tapping implicit knowledge of the different dimensions of situational awareness. This approach is similar to the one that we are currently using at the Aviation Research Laboratory to address spatial awareness in helicopter flight (Wickens, et al., submitted). At the other extreme of the measurement continuum are procedures to capture the pilot's mental model. Such an approach is currently being developed in our program of research that examines expert pilot judgment (Wickens et al., 1987)

1. **Capturing the Mental Model.** At the implicit end of the continuum, mental models may be captured by tapping knowledge structures, represented as domain-specific procedural knowledge. The psychological construct underlying these knowledge representations is that of the mental model. A mental model has been defined as a dynamic mental representation of a physical domain that aids the user in explaining and predicting system behavior (Kramer & Schumacher, 1987). Clearly this definition combines aspects of the relationships between 'what,' 'where' and 'who,' driven by an underlying temporal function.

Two global themes characterize the concepts of situational awareness and the mental model. First, each construct serves as a rubric for the structural organization of knowledge. Second, along with the representation of knowledge comes certain implications regarding expectations, interpretations, actions, and consequences for any given situation. In this respect, a pilot who is 'situationally aware' could also be described as possessing an accurate mental model of the situation. This model must be updated continuously in oder to maintain situational awareness in the dynamically changing environment.

The representation of situational awareness, as depicted in Figures 1 and 2, portrays the relational interactions between 'who,' 'what,' 'where,' and 'when.' Therefore, one may infer that the organization of this information in memory has a critical impact on the construction of an accurate situational representation, and on the resulting performance. Capitalizing on the importance of structural memory in mental models, Schvaneveldt and colleagues (Schvaneveldt, et al., 1985) sought to represent the organization of these memory structures using multidimensional scaling (MDS) techniques. MDS is a powerful statistical tool for extracting the latent structure of data from similarity ratings between pairs of data terms. Applying this technique to

key concepts involved in a specific fighter maneuver, Schvaneveldt et al. were able to distinguish between expert and novice fighter pilots. Accordingly, differences in the MDS solutions for experienced and novice pilots should provide a link to critical differences in the pilots' inferred situational awareness.

While research by Schvaneveldt et al. highlights the multivariate nature of situational awareness, the MDS assessment takes place in a static, rather than dynamic environment. In a research effort presently being carried out at the Aviation Research Laboratory, this notion of capturing the knowledge representation and situational awareness is applied to the dynamic, time-critical domain of aviation. Focusing specifically on pilot judgment and decision-making, our interest is in predicting performance through the assessment of processing limitations and procedural knowledge. MIDIS, a microcomputer-based decision simulator developed within our laboratory (Wickens, Stokes, Barnett, & Davis, 1987) provides a unique opportunity to assess situational awareness. MIDIS presents subjects with decision scenarios, or a series of "snapshots" in time, that are gradually unfolding over the course of the flight. Our previous research indicates that processing limitations provide a good prediction of novice judgments in this task, but not of experts. We have concluded that situational awareness of the expert pilot is maintained in terms of a mental model, that is updated as new information becomes available over the course of the flight. Using techniques similar to those proposed by Schvaneveldt et al. (1985), an individual assessment of each pilot's structural knowledge component will be assessed, and then used to predict performance during the cross-country MIDIS flight.

2. Spatial Awareness and Task Embedding. While situational awareness is clearly a multidimensional concept, equally important is the investigation of each of the component dimensions of situational awareness. Our understanding of spatial, temporal, and 'automation' awareness clearly needs to be developed. Very little reserach has addressed these issues. In an attempt to fill this knowledge gap, present efforts at the Aviation Research Laboratory are addressing spatial awareness.

Spatial awareness is knowledge of location and of the spatial relations between locations (Downs & Stea, 1977). The success of navigation depends on the pilot's ability to use this knowledge for remaining cognitively and perceptually attached to the environment. In particular, low level and Nap-of-the-Earth (NOE) helicpoter flight place a high demand on spatial awareness. Low level flight entails flying at altitudes as low as 50 to 100 feet while NOE involves flying as close to the earth's surface as vegetation and obstacles will permit (Headquarters of the United States Army, 1975). In such flight configurations the pilot can see little beyond his immediate surroundings. This restricted view, when coupled with the need to navigate unfamiliar terrain has important implications for map displays which can support spatial awareness.

The map is one of the most essential aids for remaining oriented within the environment. With recent advances in digital technology, map capabilities are being extended. For example, such technology is increasing the flexibility of switching between different frames of reference, and allowing different levels of detail and perspective to be displayed. Yet careful attention must be paid to configuring computer generated map displays. Clearly, those that are most effective at supporting spatial awareness will be

those that are compatible with the demands of the task and with perceptual processing (cf Woods, 1984).

Embedding tasks within a dynamically evolving flight scenario can provide an indication of the extent to which different map displays can support situational awareness. We have developed a helicopter simulation program, "Taskillan", that shows low-level and NOE flight (Wickens, et al., submitted). Subjects navigate through the simulation. Way-finding is assessed continuously, while performance on other tasks like distance estimation, target localization and reorienting when taken off the flight path is measured at periodic intervals throughout the flight (Harwood, 1987). These tasks depend heavily upon spatial knowledge of the environment and provide a good test of the effectiveness of different maps for supporting such knowledge.

Concluding Remarks

The two approaches described above share a commonality: they occur within dynamically evolving flight scenarios. Such dynamic contexts are necessary for accurately assessing situational awareness as the concept itself is a dynamically changing entity. Yet each approach is also uniquely different with respect to its focus along the measurement continuum. Situational awareness is clearly a multidimensional construct and our understanding of it will certainly be furthered by a multi-pronged approach.

REFERENCES

Downs, R.M. and Stea, D. (1977). _Maps in Minds: reflections on Cognitive Mapping._ New York: Harper & Row, Publishers.

Harwood, K. (1987). _Proposal to study map displays for helicopter navigation._ University of Illinois, Unpublished Manuscript.

Headquarters of the United States Army. (1975). FM1-1 Terrain Flight Manual.

Kramer, A.F. and Schumacher, R. (1987). Human-computer interaction: A brief glimpse of an emerging field. In L.Mark, J. Warm, and R, Huston (Eds.), _Ergonomics and Human Factors: Resistant Research_ NY: Springer-Verlag.

Marshak, W.P., Kuperman, G., Ramsey, E.G., & Wilson, D. (1987). Sit'l awareness in map displays. _31st Ann. Human Factors Proceedings,_ 533-535.

Schvaneveldt, R., Durso, F., Goldsmith, T., Breen, T., Cooke, N., Tucker, R., and DeMaio, J. (1985). Measuring the structure of expertise. _International Journal of Man-Machine Studies, 23,_ 699-728.

Stein, K.J. (1986). Complementary displays could provide panoramic, detailed battle area views. _Aviation Week and Space Technology,_ 111-115.

Stiffler, D.R. (1987). _Exploiting situational awareness beyond visual range_ (Tech Rept. 87-2370) Maxwell AFB, AL.

Wickens, C.D., Flach, J., Harwood, K., Kramer, A., & Lintern, G. (Submitted). TASKILLAN: An experimental environment to validate complex models of human performance relevant to aviation. Proceedings, 1988 Conference on Psychology in the Department of Defense.

Wickens, C.D., Stokes, A.F., Barnett, B., & Davis, T., Jr. (1987). _Componential analysis of pilot decision making._ Univ. of Illinois (Tech Rep. ARL-87-4/SCEEE-87-1).

Woods, D.D. (1984). Visual momentum: A concept to improve the coupling of person and computer. _Int'l Jo. of Man-Machine Studies, 21,_ 229-244.

Situation Awareness in the Virtual World:
Situation Assessment Report

Leslie A. Whitaker, Ph.D.
Gary A. Klein, Ph.D.
Klein Associates
P.O.Box 264
Yellow Springs, OH 45387

Abstract

High technology is making it possible to develop a virtual world
in which environmental cues and system status can be displayed for
the operator with lifelike realism. Situation Awareness is currently
a fashionable word to describe the set of environmental cues needed
to control a system in such a virtual world. However, we do not yet
have a clear picture of the cues which the operator should have
available to do the job. The present paper describes a tool for
determining these cues through knowledge elicitation from the
experienced operator.

The Need for Situation Awareness

Situation awareness (SA) is the currently popular term used to describe
the pilot's knowledge about his surroundings in light of his mission's goals.
Appropriate SA is essential for the pilot to be able to make the correct
decisions, often under time pressure, to successfully complete his mission.
Technology has taken flight far beyond the time when a pilot could look out of
his open cockpit or feel the machine through the seat of his pants. As
technological advances remove the crew further and further from the outside
world, they need displays to convey situational information. The virtual world
of Supercockpit is one more step in this progression of technology.

When the crew station of a combat aircraft becomes a virtual world, all
information about the actual world through which the aircraft is moving will
be obtained through displays and controls. The virtual cockpit is designed to
allow the pilot to wear the aircraft like a glove. This virtual glove must
convey to the crew member information which is needed without any extra
material to interfere with the execution of his mission. The virtual world
must to be designed to maintain the crew's "situation awareness" without
overloading their cognitive or response capabilities with information
irrelevant or detrimental to completing their mission. Situation Awareness is
a major contributor to the success of the mission.

Pilots, navigators, and weapons officers all must make decisions and act
in very complex situations, often under extreme time pressure. Dynamic tasks
place severe burdens on the operator. These can be tasks where the time
pressure is too great, the number of subtasks to be performed is too high, the
information flow is excessive, or the changes in environmental conditions and
goals are too abrupt. In order to perform accurately in all these situations

operators must be aware of dynamic, rapidly changing situations. Operators must do all of the following in order to maintain "situational awareness":

a. maintain their orientation even during emergency maneuvers or events;

b. maintain a sense of action priorities so they can know which subtasks to perform first, and which to omit;

c. integrate subtasks into a smooth flow of reactions;

d. develop an understanding of task parameters;

e. monitor cues that will indicate when the task has shifted.

The coupling of current computer technology with advances in cathode-ray tube (CRT), liquid crystal, and charge coupled display (CDD) technology radically altered the work station design task. Rather than designing isolated dedicated instruments such as fuel gauges and pressure gauges, designers now have the capability of displaying pictorial information and of fusing many types of information into one display. In fact, the CRT and other display surfaces in some ways present the designer with a "blank slate"; they allow completely new kinds of displays. These new displays can more directly present the layout of surfaces and the organization of events in the real world; information can be present in the basic organization of the display and need not be coded into letters, numbers, and conventional symbols. The problem becomes more pressing as we move to a supercockpit, where the capabilities of the aircraft have become enhanced, the technology is much more powerful, and the very nature of the cockpit is altered. Now the challenge is to ensure that we can promote more effective situational awareness to keep pace with technology.

Empirical Criteria for Situation Awareness

Despite its apparent simplicity, the concept of SA has been very difficult to pin down to an operational definition. Unfortunately, the term is being used by many people without a satisfactory means of assessing its presence or absence or even of defining exactly what it is. Little work has been done to guide the development of strategies to support situational awareness using CRT and other display technologies as interfaces. There is a need to develop and evaluate methods for using CRT displays to expand situational awareness for dynamic decision tasks. There has been some identification of methods that promise to be effective, additional methods should be identified. Guidelines which define adequate SA for the combat aircraft crew must be developed to the point that they can be applied to the design environments.

Since our establishment in 1978, Klein Associates has been studying decision making under time stress. In the last three years, this research focus has intensified, and we have performed a series of field studies to learn how experts make decisions. We have observed and studied urban fireground commanders (Klein, Calderwood, & Clinton-Cirocco, 1986; Calderwood, Crandall, & Klein, 1987), commanders for wildland firefighting (Taynor, Klein, & Thordsen, 1987), training device designers (Klein & Brezovic, 1986), chess

masters (Calderwood, Klein, & Crandall, 1987) and tank platoon leaders (Brezovic, Klein, & Thordsen, 1987).

In the following section, we have summarized two areas particularly relevant to the empirical determination of situation awareness:

o Knowledge elicitation through the Critical Decision Method.
o Situation awareness summarized in the Situation Assessment Report.
o Use of expertise in Recognition-Primed Decisions.

Critical Decision Method

Our initial studies of decision processes were carried out in the context of examining proficient performance. We developed methods for extracting experts' tacit knowledge and applied these knowledge elicitation techniques in a number of domains, including cardiopulmonary resuscitation (Klein & Klein, 1981), computer programming (Peio & Klein, 1984), data analysis (Klein, 1985a), and petro-chemical control room operation (Klein, 1985b). This work led us in turn into a series of studies more directly focused on the decision-making process as it occurs in real-world situations. Our first effort was an investigation of decision making by urban fireground commanders (FGCs) at the scene of a fire (Klein, Calderwood, & Clinton-Cirocco, 1985). Our choice of data-gathering method for this study was guided by our desire to model as closely as possible the natural decision making of FGCs, while meeting the demands of scientific rigor. The method chosen was a protocol analysis based on the FGC's retrospective reconstruction of his step-by-step decisions and commands at an actual incident. Incidents were chosen on the basis of their having present a command challenge, a criterion suggested by Flanagan's (1954) critical incident method. The fact that the reported incidents contained these non-routine decisions fulfilled two major requirements of the critical incident method--that recall of non-routine events tends to be superior to that of more routine cases, and that more complex or difficult cases will tend to reveal important aspects of expertise that would not otherwise be apparent (Flanagan, 1954).

The Critical Decision method developed for this initial study has now been further extended and refined. We have used the approach to examine the cognitive processes and strategies that surround decision making in several additional studies, including a replication-extension of the urban firefighter study (Calderwood et al., 1987) and investigations of decision making by design engineers (Klein et al., 1986), wildland fire commanders (Taynor et al., 1987) and tank platoon commanders during training exercises (Brezovic et al., 1987). This series of inter-related studies has allowed us to test and refine a pool of techniques and types of probes that elicit detailed, specific information about the conscious processes and strategies that underlie real-world decision making. Each study has improved the method by testing specific techniques and retaining those found to yield the most reliable and valuable information.

Situation Awareness

Our work has shown that concurrent analysis of options is much less important than is situational awareness, and we have attempted to capture key aspects of situational awareness using a representational method called a

Situation Awareness Record. The demands of working in naturalistic settings have required us to develop a Critical Decision (CD) methodology (MacGregor & Klein, 1987) that is an extension of Flanagan's (1954) critical incident technique. We have established a formalism for representing dynamic decisions by tracing the evolution of the decision maker's situation awareness as new cues and information are received.

We have recently completed an analysis of Command and Control planning sessions at Ft. Hood (Thordsen et al., 1987). During a 5-hour planning session we identified a large number of segments in the decision process. When the connections among these segments were diagrammed, it was clear that over 90% of the information on which the planners based their decisions was available immediately. Additional data were not sought; regardless of their importance (e.g., updates on enemy intent). This project has taught us two critical points relevant to SA: (1) Crew members operating under time pressure will not select alternate sources of information. The primary SA information must be displayed to the operator at the time it is needed. (2) The Situation Assessment Record is a powerful means of determining what information is needed by the operator at each segment of the mission.

Conclusions

The Situation Awareness guidelines can be used as design criteria for virtual cockpits. They will specify the environmental information which the crew must maintain in order to perform their missions. The guidelines must be specific enough to permit empirical tests of crew station design. There is currently a large gap in our knowledge of aircraft crew situation awareness needs. The situation assessment report developed from CDM interviews with crewmembers provide a powerful tool for obtaining that information.

Virtual human-machine interfaces are a technology of the near future for both military and commercial products. This transition from hardwired displays and controls to the flexibility of the virtual world permits innovative answers to questions of situation awareness. The applications of these answers to systems using the new human-computer interfaces are unlimited. Now is the time to develop accurate understandings of the SA requirements for operators of these systems. The necessity to maintain situation awareness is accentuated as the operator is removed further and further from the real world. Guidelines of the criteria for maintaining the SA are the essential basis for the design and development of successful virtual displays and controls.

References

Brezovic, C. P., Klein, G. A., & Thordsen, M. (1987). Decision making in armored platoon command. Unpublished manuscript.

Calderwood, R., Klein, G. A., & Crandall, B. W. (1987). Time pressure, skill, and move quality in chess. Manuscript submitted for publication.

Calderwood, R., Crandall, B., & Klein, G. A. (1987). Expert and novice fire ground command decisions (MDA903-85-C-0327). Alexandria, VA: U.S. Army Research Institute.

Flanagan, J. C. (1954). The critical incident technique. _Psychological Bulletin_, _51_, 327-358.

Klein, G. A. (1985a). _Knowledge engineering analysis of data extraction strategies_. Technical Report prepared for Perceptronics.

Klein, G. A. (1985b). Working paper for Beville Engineering, Inc.

Klein, G. A., & Brezovic, C. P. (1986). _Human performance data needed for training device design decisions_ (F33657-82-C-0153). Wright Patterson Air Force Base, OH: Air Force Aerospace Medical Research Laboratory.

Klein, G. A., Calderwood, R., & Clinton-Cirocco, A. (1985). _Rapid decision making on the fire ground_ (KA-TR-84-41-7). Alexandria, VA: U.S. Army Research Institute.

Klein, G. A., Calderwood, R., & Clinton-Cirocco, A. (1986). Rapid decision making on the fire ground. _Proceedings of the Human Factors Society 30th Annual Meeting_, _1_, 576-580. Dayton, OH: Human Factors Society.

Klein, H. A., & Klein, G. A. (1981). _Perceptual/cognitive analysis of proficient cardio-pulmonary resuscitation (CPR) performance_. Paper presented at the Midwestern Psychological Association Meetings, Detroit, MI.

MacGregor, D., & Klein, G. A. (1987). _Knowledge elicitation of recognition-primed decision making_ (MDA903-86-C-0170). Alexandria, VA: U.S. Army Research Institute.

Peio, K. J., & Klein, H. A. (1984). Skilled performance: Computer software. _Proceedings of the 28th Annual Human Factors Society Conference_. Dayton, OH: Human Factors.

Taynor, J., Klein, G. A., & Thordsen, M. (1987). _Distributive decision making in wildland forest fighting_. Unpublished manuscript.

What's Up? A New Orientation

Gary E. Riccio
Armstrong Aerospace Medical Research Laboratory
Wright-Patterson AFB, OH

Eric J. Martin
Systems Research Laboratories, Inc
Dayton, OH

Thomas A. Stoffregen
At Large

In this paper we describe the results from an experiment on the perception of upright. Subjects were restrained in a device that rotated about an anterior-posterior axis through the buttocks. Subjects estimated their mean tilt with respect to upright after controlling their roll orientation for 31 seconds in the presence of a continuous roll-axis disturbance. The dynamics of the device simulated an inverted pendulum with a direction of balance that was varied randomly. We evaluated the relative effects of gravity and the balance point on subjective estimates of tilt. The results indicate that the balance point had a significant effect on the perception of upright. This finding has important implications for theories of orientation.

Maintaining the orientation of our bodies is fundamental to behavior (Gibson, 1966; Riccio & Stoffregen, 1988). Orientation is fundamental because of the constraints, on interactions with the environment, imposed by the configuration and stability of the body. For example, the configuration of the body partially determines what aspects of the environment are accessible to perception and action systems. In addition, the stability of the body (or lack thereof) limits the precision of perception and action. Stoffregen & Riccio (1988) have developed a new theory of orientation that is based on the interaction of an organism with its environment.

> The act of orienting...is controlled interaction between an organism and a surface (or surfaces or medium) of support, so as to maintain dynamic equilibrium with respect to the forces acting on the organism. Both force and some surface to resist the force are necessary - neither is sufficient alone. Classical definitions of orientation assume the comparison of body axes to some external reference frame (Howard & Templeton, 1966; Schone, 1984). In contrast, our definition does not involve such a comparison but instead depends on the attainment of a particular kinematic state. (Stoffregen & Riccio, 1988, p.5).

Classical approaches to orientation are based on the assumption that gravitoinertial force (the sum of gravitational and inertial forces) is perceived. However, Stoffregen & Riccio (1988) review several experiments that contradict this assumption. These experiments tested the ability of humans to indicate the direction of "vertical" with eyes closed while under water; typical errors were on the order of 20 degrees. Stoffregen & Riccio (1988) argue that "if the phenomenal upright correlates with anything, it should correlate with the goal of the act of orienting" (p. 7), and in their view, the goal of orientation is to attain a particular kinematic state. This may simply involve orienting in the direction for which the body is most stable or is "balanced" with respect to external forces (Riccio & Stoffregen, 1988). In the underwater experiments, the force of gravity had no influence on the kinematics (movements) of the body and there were no direction-specific constraints on the control of body movements. The results of these experiments are, therefore, consistent with a kinematically based theory of orientation, but they provide no evidence that kinematics can be the basis for the perception of orientation.

We have devised a method for manipulating the direction of balance for the body independently from the direction of gravity. This methodology allows us to experimentally contrast our approach with the classical approach to orientation. In the present experiment, the two approaches lead to different predictions about the perceived "upright" that are, unlike in the underwater experiments, distinguishable from the null hypothesis.

Methods

Subjects

Subjects were paid volunteers recruited from local universities. Nine subjects have completed the experiment to date. One subject did not complete the experiment because he was unable to control the experimental device. The re were two groups of subjects. The first group consisted of five males, and the second group consisted of three females and one male (see *Procedure*).

Apparatus

Subjects sat in a device that rotated about an anterior-posterior axis passing through the buttocks. The subjects wore helmets, as well as, shoulder and lap restraints typical of high-performance fighter aircraft. The restraints limited the movements of the torso to the "roll" axis. In one experimental condition, subjects wore a soft cervical collar that reduced movements of the head relative to the torso and the experimental device. The dynamics of the device were based on a simple inverted-pendulum model of human stance (cf. Stoffregen & Riccio, 1988). Accordingly, the device was unstable such that, without the subject's control, its angular acceleration would be proportional to its angular position (tilt) with respect to a preselected position or "balance point" (Figure 1). Subjects could control the roll rate of the device by applying force to an isometric stick located on the right armrest. The balance point could be set at various angles with respect to the direction of gravity.
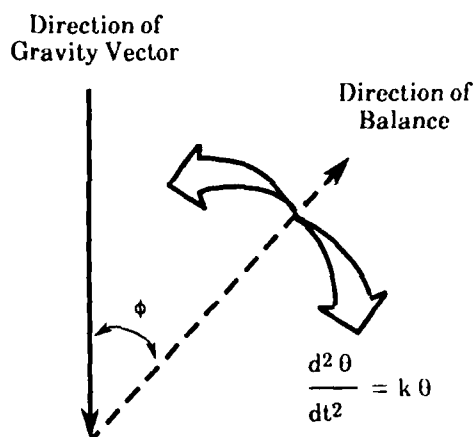
Figure 1. Manipulation of the direction of the balance point independently from gravity. ϕ is the angle between the balance point and the gravity vector, and θ is the instantaneous deviation from the balance point.

Procedure

Task. The subjects' task was to keep the device "upright" while the device was perturbed by a continuous pseudo-random wide-bandwidth disturbance that was equivalent to 0.371 newton meters of root-mean-square torque on the control stick. They were also instructed to minimize the motion of the device. Each trial was approximately 31 seconds in duration. Each trial began with the device, the subject, and the balance point aligned with the direction of gravity so that they would not be given advanced knowledge about the tilt of the balance point. During the first seven seconds of the trial, the balance point was ramped to a preselected location; this was done so that the subjects would not start a trial at an uncontrollable tilt with respect to the balance point. After each trial, subjects estimated the direction and magnitude of their average tilt with respect to upright. No subject had

any problems with the instructions, and no details were given about what was to be the reference for their estimates.

Training. There was one hour of training (on day one) in the task for each subject. The training consisted of a series of trials with dynamics that were progressively more difficult to control. The difficulty was operationally defined by the constant of proportionality between angular position and angular acceleration (Figure 1). There were four steps in the training sequence: defined by constants of 0.5, 1.0, 1.5, and 2.0 s^{-2}. Subjects transitioned to the next step after completing five consecutive trials without "falling over". The fourth step was with the dynamics used in the experiment proper. Throughout training the balance point was aligned with gravity.

Experimental Design. On days two through seven, subjects participated in the 168 trials of the experiment proper (28 trials per day). There were 24 trials for each of seven settings of the balance point: -15, -10, -5, 0, +5, +10, and +15 degrees with respect to gravity. The variation in angle between the balance point and gravity increased the sensitivity of the design by reducing the correlation between tilts with respect to the balance point and gravity. The angle between the the balance point and gravity was randomized over each of two sets of 84 trials. In one set of 84 trials, subjects wore the cervical collar; the collar was not worn in the other set. Nine subjects have completed the experiment to date. The first group (five subjects) wore the collar in the first set, and the second group (four subjects) wore the collar in the second set. (Each group will be completed with a total of six subjects.)

Results

Data Reduction

For each subject and collar condition, the "subjective" estimates of mean tilt with respect to upright were correlated with the "objective" mean tilt with respect to gravity and with respect to the balance point. Correlations were computed separately for direction only and magnitude only. The phi statistic was used to compute the correlations for direction. The Pearson product-moment correlation coefficient was computed on rank scores for the magnitude data. The correlations for magnitude were computed only on those data for which the direction was correctly estimated. These correlations provide complementary information. Correlations were computed separately for direction and magnitude because the correlations between the tilts with respect to gravity and the balance point were lower than if signed magnitudes were used. Since the tilts with respect to gravity and the balance point are correlated, their independent effect on subjective tilt (with respect to upright) can only be assessed with partial correlations. Partial correlations for each subject were derived from the simple correlations, for direction only and magnitude only, using the following formula.

$$R_{ub.g} = \frac{R_{ub} - R_{ug} R_{bg}}{(1 - R_{ug}^2)^{.5}(1 - R_{ub}^2)^{.5}} \qquad R_{ug.b} = \frac{R_{ug} - R_{ub} R_{bg}}{(1 - R_{ug}^2)^{.5}(1 - R_{ub}^2)^{.5}}$$

$R_{ub.g}$ is the partial correlation between the subjective tilt with respect to upright (u) and the objective tilt with respect to the balance point (b), controlling for the effects of gravity (g). $R_{ug.b}$ is the comparable partial correlation with respect to gravity, controlling for the balance point. The correlations on the right-hand side of each equation are simple correlations (either phi or Pearson coefficients) involving the balance point, gravity, and subjective upright. Inferential statistics were *not* computed for individual correlations; they were used simply as "index numbers" (i.e., basic data) for each subject. All inferential analyses were computed over these index numbers, that is, over subjects.

Table 1. Simple and partial correlations between subjective and objective tilts.

		Direction			Magnitude		
		Mean	T	P	Mean	T	P
Simple	R_{ub}	.502	10.4	.0001	.281	18.1	.0001
Correlations	R_{ug}	.436	11.2	.0001	.320	7.7	.0001
	R_{bg}	.637	13.0	.0001	.516	16.0	.0001
Partial	$R_{ub.g}$	.297	5.0	.0010	.142	5.1	.0009
Correlations	$R_{ug.b}$	.152	5.1	.0009	.192	4.2	.0030

Inferential Analyses

For each correlation, or difference between correlations, we used two-tailed t tests (df = 8) to evaluate whether the mean over the nine subjects was significantly different than zero. The difference between the two collar-conditions was not significant (p > .05) with respect to any of the simple or partial correlations. The statistical tests reported below were performed on the means of the two collar-conditions from each subject. The correlations and the associated statistics are presented in Table 1; all correlations are significantly greater than zero (p < .01). The means and the confidence intervals for the partial correlations are depicted in Figure 2. The difference between the partial correlations with respect to gravity and the balance point were also evaluated (Figure 2). The difference between these correlations (balance point minus gravity), when computed on direction only, was marginally significant (mean = .145, t = 2.135, p = .065). When computed on magnitude only, the difference between the correlations was negligable (mean = -.051, t = -.0802, p = .446).
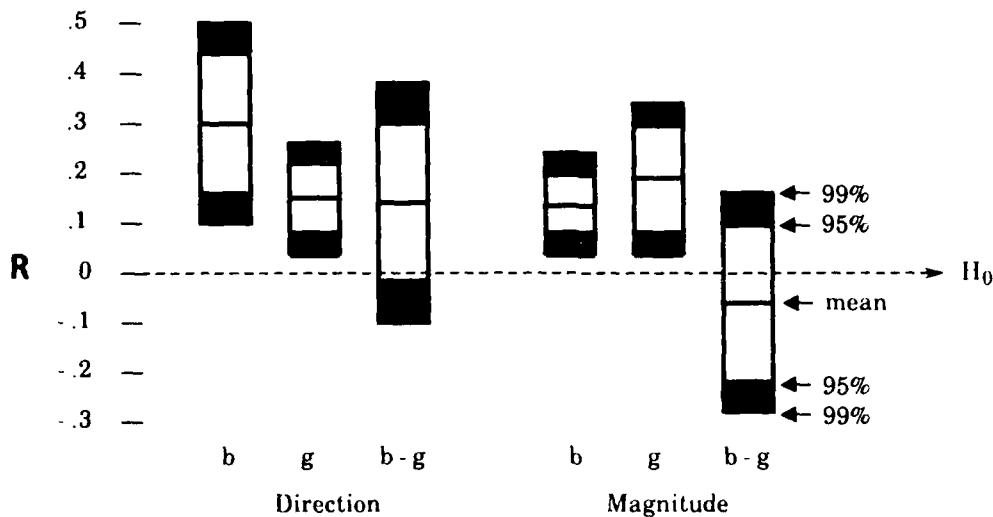

Figure 2. Correlations between subjective deviation from "upright" and objective deviation from the balance point (b) or gravity (g), and difference between correlations (b - g). Bars depict the means and 95% and 99% confidence intervals for partial correlations computed on direction only and magnitude only. See text for details.

Discussion

The magnitude of the correlations between subjective and objective tilts indicates that the subjects performed quite well in the estimation task; that is, their estimates were reliably related to objectively defined references (the balance point and gravity). Summing the appropriate pairs of squared simple and partial correlations for direction-only and magnitude-only provides a rough estimate of the amount of variance in the ratings due to the variation in tilt relative to the balance point and gravity. By this estimate, 40% of the variance is accounted for. This indicates that we were successful in developing a paradigm for studying the perceived upright in a situation where the goal of the act of orienting was well defined (cf. Stoffregen & Riccio, 1988, p.10).

The most important finding of this study is that *a kinematic property, the balance point, is a determinant of the perceived upright.* This radically new finding has important implications for theories and experimentation on orientation. For example, rather than comparing some poorly defined body axis to an external axis (e.g., gravity) that is independent of the body, it seems that we are directly sensitive to, and control, the interaction between our bodies and the environment. This would obviate the need for the awkward concept of "sensory conflict" between external references (e.g., between gravitoinertial force and anisotropic optical structure) which would have to apply to any situation in which an organism or the environment moves (Riccio & Stoffregen, 1988; Stoffregen & Riccio, 1988). In our view, perceptual systems function cooperatively to pick up redundant or complementary information rather than antagonistically to resolve conflict. Our findings also suggest that kinematics (patterns of movement) are a potentially rich source of information, about the characteristics of an organism and its enviroment, that should be studied by perceptual psychologists (Carello, Turvey, & Kugler, 1985; Riccio & Stoffregen, 1988; Stoffregen & Riccio, 1988).

The results from this experiment indicate that both gravity and the balance point influence the perception of upright (although there are also indications that the effect of the balance point may be greater). This is not neccessarily due to perception of gravity. We could not completely eliminate the effects of gravity on the movements of the subject; for example, the movements of the head, arms, and legs were influenced both by gravity and by the motion of the experimental device. We will attempt to minimize these confounds in future experiments. It is also possible that the spatial patterns of pressure on the body (due to seat restraints, arm rests, etc.) provide a sensitivity to gravity that is greater than that indicated by the underwater experiments. This will also be investigated in future experiments.

There is a potential application of these findings to flight simulators. We were able to reduce the impression of tilt when subjects were titled with respect to gravity. A severe limitation of "moving-base" flight simulators is that tilt and rotation cannot be independently manipulated; for example, unlike the experience of a "coordinated turn" in flight, one feels tilted after a simulator rotates. In principle, one could design a simulator that pushes and pulls on the pilot in such a way that the balance point(s) for the body could be manipulated independently from the pilot's orientation relative to external references such as gravity.

References

Carello, C., Turvey, M.T., & Kugler, P.N. (1985). The informational support for upright stance. *The Behavioral and Brain Sciences, 8,* 151-152.

Gibson, J.J. (1966). *The senses considered as perceptual systems.* Boston: Houghton-Mifflin.

Howard, I.P. (1966). *Human spatial orientation.* New York: Wiley.

Riccio, G.E. & Stoffregen, T.A. (1988). Affordances as constraints on the control of stance. *Human Movement Science,* in press.

Schone, H. (1984). *Spatial orientation: The spatial control of behavior in animals and man.* (C. Strausfeld, Trans.). Princeton, NJ: Princeton University Press.

Stoffregen, T.A. & Riccio, G.E. (1988). An ecological theory of orientation and the vestibular system. *Psychological Review, 95,* 1-11.

Acknowledgements

Data Base Requirements for Simulating High Resolution Radar

Peter M. Crane, Ph.D.
Air Force Human Resources Laboratory
Operations Training Division

Abstract

Simulation of ground mapping radar requires a data base describing the location and characteristics of features on the ground. Pilot studies indicated that the amount of detail required to support simulation of high resolution radars would exceed production capacity. Behavioral research studies were conducted to: (1) determine the task critical features required to effectively use the radar, and (2) develop and test alternative data bases.

The major cost driver in simulating high resolution ground mapping radar such as the Synthetic Aperture Radars (SAR) employed on the F-15E and B-1B is the generation of the feature analysis data base. This data base consists of descriptions of objects and features on the surface of the earth such as roads, rivers, buildings, and areas of vegetation. Increasing the density of this data base improves the fidelity of SAR simulations but requires significant increases in cost and production time. The density of this data base is largely determined by its nominal capture criterion, i.e., the smallest horizontal dimension of a feature for individual portrayal. Objects smaller than the capture criterion are either not included or are aggregated with other objects and portrayed as an area of uniform reflectivity. A prototype feature data base with a 10m capture criterion was found to produce acceptable SAR simulations. Production in quantity, however, would quickly overwhelm the available resources.

Behavioral research methods can be used to identify the degree of fidelity required to fulfill simulation requirements. Simulations of varying levels of fidelity can be prepared for evaluation by subject matter experts. Crane and Bell (1985) used the method of paired comparison with Thurstone Scaling analysis to determine the maximum perceptible level of transformation accuracy in simulated real beam ground mapping radar imagery. Simulations were produced at six levels of transformation accuracy and then evaluated by KC-135 and C-130 navigators. It was found that increases in accuracy beyond a certain point were not detectable by trained radar operators.

Similar studies have been conducted on the data base requirements for simulating SAR. In a preliminary experiment, SAR experienced radar operators were asked to examine simulations produced using data bases with 10m, 15m, 20m, and 30m capture criteria and to specify, "the minimum level of image quality required to simulate SAR for any purpose." It was found that a 15m capture criterion was adequate. However, the change from 10m to 15m would not significantly affect cost or production. Based on this finding, it was concluded that a change in the capture criterion alone would not be sufficient to develop a cost efficient data base. Instead, it was hypothesized that a

lesser degree of ground truth information could be combined with synthetic enhancements without reducing simulation effectiveness. Therefore, studies were conducted to determine: (1) what types of features must be portrayed accurately to perform SAR tasks, and (2) how the level of performance of B-1B Offensive Systems Officers (OSOs) would be affected by using SAR simulations generated from different types of data bases.

Radar Scope Interpretation Cue Analysis

Radar Scope Interpretation (RSI) cues, or pointers, are reflectors used by radar operators to aid in locating an aimpoint. These surface features are identified from charts or photographs during pre-mission study. Typically, the operator will select distinctive features within the area to provide overall orientation and will then locate successively smaller features to help locate the target. The data collected in the preliminary study suggested that not all features within a scene are of equal importance to SAR operators. The major difference between a 10m and 15m data base is the depiction of features the size of houses. These features are individually depicted in the 10m data base while in the 15m data base houses within residential areas are aggregated into a single feature. SAR operators found 10m and 15m data bases equally acceptable provided that: (1) the aimpoint was not located within high density areas, and (2) the road patterns in high density areas were visible. This finding suggests that the houses themselves were not task critical features but that the road pattern was task critical. The RSI cue analysis was conducted to more systematically determine what features are used by SAR operators in performing their tasks.

Research Methods

In premission study, B-1B OSOs examine Fixpoint Graphic cards which contain the coordinates of the specified aimpoint and a 1.25 nm x 1.25 nm overhead photograph of the area centered on the aimpoint. The OSO will use the Fixpoint Graphic to identify RSI cues on the photograph. For the RSI analysis, unclassified Fixpoint Graphic cards for 115 aimpoints within the Strategic Training Range Complex were assigned to one of four categories based on the specified aimpoint and the surrounding features. The categories were: Urban, Small Group, Isolated, and Terrain. The Fixpoint Graphic cards were divided into five sets of 23 with approximately equal numbers of the four types in each set. OSOs were asked to examine the Fixpoint Graphic cards in one of the sets and to highlight the RSI cues they would expect to see on the B-1 SAR; 22 OSOs from the 96th BMW, Dyess AFB, TX participated.

Results and Discussion

Each highlighted feature was assigned to one of four pointer classifications: Lines of Communication (LOCs) such as roads, railroad tracks, or power lines; Natural Features including vegetation, rivers, lakes, treelines, or plowed fields; Structures defined as man made objects occupying three dimensional space; and Terrain. Overall, 56% of the highlighted pointers are LOCs. These are linear features which extend across much of the scene. Although intersections and grade crossings were not scored as separate features, the OSOs often commented that they would typically locate the major LOCs, follow them to an intersection and then locate the aimpoint relative to

the intersection. Structures were also frequently highlighted as pointers within Urban and Small Group scenes. The structures typically highlighted were the largest or tallest buildings within an area such as a tower or grain elevator. Natural Features were the most common pointers within Isolated scenes. Since Isolated scenes were defined by the absence of surrounding man-made objects this result is not unexpected. Terrain was used as a pointer only where the aimpoint itself was a terrain feature.

The data on RSI cues provided by SAR operators revealed that the density of ground truth detail provided by 10m and 15m data bases may not be required to perform SAR tasks. Ground truth information is required for LOCs, major buildings, and significant natural features; these data are provided in a 30m data base. The major problem with the 30m data base is the aggregation of clusters of objects such as houses into areas of uniform reflectivity. However, since the individual structures are not used as pointers, these details might be replaced with generic information such as a texture pattern added to the simulation. An enhanced 30m data base would then closely resemble an actual SAR image and would allow the operator to use a cluster of small buildings as a pointer even though the information about the individual buildings would not be accurate.

SAR Task Performance Evaluation

When capture criterion alone was varied in the preliminary study it was found that a 30m data base was rated unacceptable. Based on the RSI cue analysis however, it was hypothesized that a 30m data base could be enhanced using low cost procedures and would support B-1B OSO task performance. To test this prediction a simulator based study was designed using performance on a navigation update task as a measure of simulation utility.

Research Methods

Simulated SAR images were generated from 10m and 15m data bases and from enhanced versions of 30m and 100m data bases. The four data bases plus actual SAR simulations were obtained for 15 areas within the Strategic Training Range Complex and were recorded for presentation in a B-1 simulator. The 15 scenes were blocked into three groups: Urban, Small Group, and Isolated. Each subject saw only one image for each scene and the combinations of scene and image type were balanced across subjects so that subjects saw three scenes from each data base and three scenes of actual SAR. The dependent variables were deviation in crosshair placement from the aimpoint and confidence in placement accuracy. The subjects were 25 OSOs from the 96th BMW who had a median of 220 hours of B-1B time and 1000+ hours total time.

Each subject was given ten minutes to study the 15 Fixpoint Graphic cards and then entered the simulator. The subject's task was to locate each aimpoint as accurately as possible within 60 seconds. A trial began when the OSO was cued to request a SAR image. He then compared the Fixpoint Graphic to the image and moved the crosshairs to the target. When he was satisfied with the placement he designated by pressing a button and then indicated his confidence on a 7-point scale with the anchors, "1-Complete Guess," to, "7-Very High Confidence." After performance data were collected the OSOs were shown photographs of the images they had seen and asked to rate their acceptability for Mission Rehearsal.

Results

The crosshair placement errors for each scene type and data base were analyzed using an incomplete blocks model analysis of variance. This analysis showed that data base significantly affected accuracy for Urban [$F(4, 90)=3.49$, $p=.011$] and Small Group [$F(4, 91)=3.10$, $p=.019$] fixpoints; no effect was found for data base with Isolated fixpoints ($p>.05$). The significant main effects were further analyzed using Fisher's Least Significant Difference test. This analysis showed that the placements using the 100m enhanced data base were significantly less accurate than the 30m enhanced, 15m, and 10m data bases and actual SAR for both Urban and Small Group scenes. The placement accuracies for 30m enhanced, 15m, and 10m data bases and the actual SAR were not significantly different from each other. The confidence ratings were also analyzed using an incomplete blocks analysis of variance. There was a significant effect due to data base for Urban [$F(4,82)=9.18$, $p<.0005$], Small Group [$F(4, 90)=8.59$, $p<.0005$], and Isolated scenes [$F(4, 90)=4.09$, $p<.004$]. Confidence was lowest for the enhanced 100m data base for all scene types. There were no significant differences between the enhanced 30m, 15m, and 10m data bases or the actual SAR. There were also no significant differences on the ratings for acceptability for mission rehearsal between the enhanced 30m, 15m, and 10m data bases; the actual SAR, however was rated as more acceptable for mission rehearsal than any of the simulations.

Discussion

The SAR Performance Evaluation study was based on the prediction that a 30m data base enhanced with generic information about high density areas would support task performance as well as data bases containing all ground truth information. This prediction was confirmed; there were no significant differences on crosshair placement accuracy, operator confidence ratings, or ratings of acceptability for mission rehearsal between 10m, 15m, or enhanced 30m data bases. Adding enhancements to a 100m data base did not produce usable SAR simulations.

Conclusions

The results of the preliminary investigation indicated that 15m was the minimum acceptable capture criterion for generating acceptable SAR simulations. In contrast, the RSI analysis showed that small features were not used as pointers or aimpoints were therefore not task critical. SAR operators however needed to be able to distinguish between high density areas and areas of true uniform reflectivity such as a forest or cultivated field. Adding appropriate enhancements to the 30m data base provided this task critical information at significantly lower cost than collecting ground truth information about each structure in the scene.

References

Crane, P. M. & Bell, H. H. (1985). Flight training simulators: Effects of terrain accuracy on simulated radar image quality. (AFHRL-TP-85-28) Brooks AFB, TX: Air Force Human Resources Laboratory.

Stress Resistant Displays

Leslie A. Whitaker, Ph.D.
Gary A. Klein, Ph.D.
Klein Associates
P.O.Box 264
Yellow Springs, OH 45387

Abstract

System operators can often overcome poorly designed display panels and incompatible display-control relationships during routine system operation. However, when time-pressured decisions and responses must be made, the display designs must be optimal if errors are to be avoided. In the present paper, we describe stress resistant display guidelines related to five topics: metaphors, streamlining, functional prototypes, foregrounding, and fusion.

Klein Associates reported the effects of displays on time pressured responses to industrial emergencies (Klein & John, 1985). We found that most displays are optimized to handle the 95%+ of cases that are routine, and they are not well suited for the 1-5% of cases that are marked by extreme time pressure (i.e., emergencies). This paper discusses display principles for addressing those rare but critical cases. Five areas of display guidelines were selected to illustrate the power of current high technology hardware (CRTs, liquid crystals, and charge coupled displays).

Display Guidelines

Metaphors

Lakoff and Johnson (1980) defined metaphor as the use of an example in one domain to provide structure for a second domain. For example, they explained that there is not a clear concept of what an argument is. But if we say that ARGUMENTS ARE LIKE WAR, then we can use what we know about war to see how our opponent is attacking our position, and we are defending it and searching for chances to counterattack. In a study of designers (Dent, Klein, & Eggleston, 1987), we found that metaphors were employed in virtually all the displays we examined.

Metaphor is a potentially powerful source of organization in CRT displays in two ways: to the designer as a source of organization in guiding decisions about how to portray information, and to the user in guiding attention to important information needed for skilled action under time pressure. Metaphor is powerful because it uses what is well known and familiar to comment on or depict what is less well known.

With an increase in the use of pictorial displays comes more opportunity for visual metaphor. Indeed, metaphors are pervasive in designs for interfaces in the areas of word processing and animation (Carroll & Mack, 1985). Organizing metaphors which structure a whole display or set of

displays, and visual metaphors which can appear in iconic displays, seem to be important tools on which designers of word-processing software and interfaces can draw.

The resemblance between different domains is the basis of the power of metaphor to guide attention to information about real-world objects and events (Verbrugge & McCarrell, 1977; Verbrugge, 1980; Dent et al., 1987). This resemblance, then, supports the transfer of skilled action known well in one domain to action in another domain. The effective metaphors were ones that guide performance by letting the operator access an integrated response sequence from another domain and use it to react to the new domain. The function of metaphor was to guide actions, not to structure the assessment of states.

When used systematically, metaphors can be very powerful in organizing the designer's task and in organizing the operator's use of the display; examples are the flying-in-formation metaphor that results in a phantom wing leader portrayed for the pilot and the desktop metaphor that organizes commands and icons for word processing.

The power of metaphor lies in its potential to organize displays and guide the operator's interactions with the displays. Domains that are well known to both designer and user can be used as metaphors to coordinate displays and the actions the displays support. This is why the flying-as-driving metaphor and the word-processing-as-typing metaphors work so well. This is also why the "health" metaphor for the status of an aircraft might not work; physiology and diagnosis are not well known or often used by designers and pilots. Furthermore, the health metaphor does nothing to guide the operator's reactions.

Because two different domains are involved in metaphor, mismatches or areas of dissimilarity will exist and so the potential to hide certain information also exists. The challenge is to develop guidelines and support material to maximize the effective use of metaphor by designers and to minimize the risk that metaphor could mislead the operator. The cost of the misuse of metaphor may be high; however, the cost of not using metaphor may also be high. The designs that will be used in cockpits sometime in the next 15 years already include metaphors, but they have not been used as consistently or as completely as possible. In addition, alternative metaphors or alternative designs without metaphor have not been tested against those already in use.

Streamlining

Eliminating nonessential elements from a display is known as streamlining. This can include intelligent decluttering. Under time pressure, the operator needs to find the critical cues with the least amount of distraction. Because of the ability to add more and more detail to displays, designers are currently preparing strategies to "declutter" a display by reducing the classes of details. However, this can take time. The operator must take time out to call for decluttering. What is needed is a set of techniques that would produce the decluttering as a function of mission requirements, unless over-ridden by the operator.

Functional prototypes

This means an orientation on the person's preparations for action rather than on the passive identification of the cues in a scene. In our fireground research, standard structural prototypes for fires would include residences, apartment buildings, factories, etc. We did not find much evidence for the use of these structural prototypes. The fireground commanders were oriented around the orders they would have to give, and the prototypes were things like Search & Rescue operations, interior attacks, defensive operations to contain the fire, etc. The commonality was not in the structural features but in the functional requirements of the fires. Displays that presented functional (response) layouts will be easier to use during time pressure.

The display organization can be focused on reactions that the operator must make, rather than on simply producing status descriptions. An examples is the use of status descriptions for process controllers in a chemical plant. Each CRT screen portrays a separate chemical tank. Unfortunately, during emergencies the operator needs only a small amount of information from each screen and must flip back and forth between them in order to have the information necessary to achieve a rapid and safe shutdown. Ideally, there would be functional screens for the major operations of shut-down, start up, and damage control.

Bateman, Reising, Herron, and Calhoun (1978) have shown the effectiveness of a multifunctional keyboard. They demonstrated that a tailored display had definite performance gains over the standard logic tree. Cockpits cannot present dedicated switches for each possible function; if they did there would be so many switches that there would be little room for anything else. Bateman et al. used some dedicated switches and some multifunction switches. In some conditions, the multifunction switches changed functions according to a branching logic. Under another condition, there was an automatic assignment of functions and legends to switches according to the flight mode of the aircraft. Significant time savings were realized using the tailored logic, which would be predicted by our guideline of functional prototypes.

Foregrounding

Highlighting the key aspects of a task helps to set task priorities within a display. Figure/ground relationships can help to emphasize the key aspects of the task. We have found that even under routine operations it can be difficult to maintain a proper sense of priorities, to ensure that the most important task is done first. Under time pressure, this need is even greater. Operators can become so immersed in handling noncritical problems that they fail to react to emergency conditions until it is too late. This was the rationale for alarms, but auditory alarms are not the sole source of alerting available to the designer. There is a need to restructure displays to aid the operator in setting task priorities.

For example, during extreme conditions such as spins, the F-14 CRT will degrade into a large arrow showing the pilot which way to pull the yoke. There is a general requirement to help the operator focus on the most critical task first. As a second example, instances of figure/ground

337

reversals should be considered here. Pilots need to see where the radars of anti-aircraft batteries are searching, but when these become too extensive there is a need to shift into a display of safe routes through the anti-aircraft irradiating spotlights.

Fusion

Subtasks can be combined to reduce workload. It is sometimes possible to synthesize several tasks. This is especially true where different requirements groups have each generated a set of subtasks and the display designer includes them all into the design, without trying to determine the overlaps and opportunities to simplify workload by changing task concepts and blending subtasks together. The work of Hirst, Spelke, Reaves, Caharack, and Neisser (1980) has shown how effectively subjects can fuse totally disparate tasks, and in so doing evade the constraints of workload accumulation. In a sense, a beginning pilot is performing a large set of discrete tasks until practice allows them to be fused together; at that point, workload decreases and performance becomes smoother. One goal of display design would be to support the more rapid fusion of tasks and subtasks.

Conclusions

Further research must be conducted to determine the limits of these display guidelines. They are not yet refined to the point that designers can use them in handbook fashion. Instead, we hope that guidelines such as these will focus research on the special needs of operators making time-pressured decisions. They need all the help we can give them.

References

Bateman, R. P., Reising, J. M., Herron, E. L., and Calhoun, G. L. (1978). Multifunction Keyboard Implementation Study (F33615-76C-0013). Wright Patterson Air Force Base, OH: Air Force Flight Dynamics Laboratory.

Carroll, J. M., & Mack, R. (1985). Metaphor, computing systems, and active learning. International Journal of Man-Machine Studies, 22, 39-57.

Dent, C., Klein, G. A., & Eggleston, R. (1987). Metaphor casting of information display requirements (KATR-864-87-01Z). Yellow Springs, OH: Klein Associates Inc. Prepared under contract F33615-86-C-0544 for Aeronautical Systems Div/PMRSB, Wright Patterson Air Force Base, OH.

Hirst, W., Spelke, E. S., Reaves, C. C., Caharack, G. & Neisser, U. (1980). Dividing attention without alternation or automaticity. Journal of Experimental Psychology: General, 1, 98-117.

Klein, G. A., & John, P. G. (1985, August). Predicting workload requirements during the predesign stage. Paper prepared by Klein Associates for Armstrong Aerospace Medical Research Laboratory, Wright-Patterson Air Force Base, OH.

Lakoff, G., & Johnson, M. (1980). <u>Metaphors We Live By</u>. Chicago, IL: University of Chicago Press.

Verbrugge, R. (1980). Transformation in knowing: A realist view of metaphor. In R. Honeck & R. Hoffman (Eds.), <u>Cognition and figurative language</u>. Hillsdale, NJ: Lawrence Erlbaum Associates.

Verbrugge, R., & McCarrell, N. (1977). Metaphoric comprehension: Studies in reminding and resembling. <u>Cognitive Psychology</u>, <u>9</u>, 494-533.

Simulator Design Features for Precision Helicopter Hover over Small Ships

Daniel J. Sheppard*, Sherrie A. Jones*,
Joyce Madden**, and Daniel P. Westra*

*Essex Corporation
**Naval Training Systems Center

Abstract

The research was designed to evaluate the effects of some current design features of the SH-60B Operational Flight Trainer (OFT) and proposed simulator design modifications on pilot performance in a precision hover task. These were: (1) scene detail (SH-60B OFT scene versus an upgraded scene), (2) field-of-view (wide versus a smaller SH-60B OFT field-of-view), (3) visual display lag (117 msec versus 183 msec), (4) dynamic seat cueing (on versus off), and (5) dynamic inflow (standard rotor model available in existing trainers versus an updated rotor model). On the basis of the factors studied in the experiment, the wider field-of-view, the shorter visual display lag, and the upgraded rotor model are recommended for use. The dynamic seat cueing as evaluated in this experiment is not recommended. Scene detail did not have a significant effect on performance.

The Visual Technology Research Simulator (VTRS) at the Naval Training Systems Center (NTSC) has been experimentally evaluating simulator design options and training procedures for a wide variety of flight tasks. A current VTRS research effort has focused on the evaluation of design features for a training simulator to support the acquisition of skills needed to execute helicopter landings on small ships. A Vertical Takeoff and Landing (VTOL) simulator, which simulates the Navy's SH-60B Seahawk helicopter, is used to examine these issues. The present experiment investigated the effects of current design features of the SH-60B Operational Flight Trainer (OFT) and some proposed design modifications on pilot performance in a precision hover task. The hover segment is considered to be among the more critical components of helicopter shipboard landing.

Method

Experimental Issues

Scene Detail. Pilots have indicated that they attend to specific visual features while monitoring their position and orientation relative to the ship and sea from the initial phase of tracking the ship to the final phase of touchdown (Berbaum & Kennedy, 1985). Many of these features consist of markings and reference points located on the hangar wall or ship's deck which convey very distinct types of spatial information. The manipulation of scene content employed in the current study involved a comparison of performance obtained with two dusk scenes, one of which is representative of the visual scene (FFG-7) currently available in the SH-60B OFT, versus an upgraded version of the FFG-7 available at VTRS (additional deck markings).

Dynamic Seat Cueing. G-seats represent a relatively new technology in helicopter simulation. Consequently, few studies with helicopter g-seat cueing have been performed and results concerning their utility for simulation have been inconclusive. G-seat cueing was examined in a previous helicopter shipboard landing experiment at VTRS (Westra & Lintern, 1985). The g-seat effects were not significant and pilots' reported problems with

the acceleration cues given. A new algorithm based on the work of McMillan, Martin, Flach, and Riccio (1985) emphasizing position cueing in the vertical dimension was developed at VTRS and used in the experiment.

Field of View. The task of determining an optimal, cost-effective visual system field of view (FOV) for the simulation of helicopter shipboard landing remains unresolved. The demands of the operational environment, particularly during the hover and landing segments, require that the pilot be provided with an adequate lower FOV. A pilot's attention in these two phases of the task is largely directed toward acquiring visual information from the surface of the ship's hangar wall and deck through the chin window(s) and lower portion of the forward and side windows. Figure 1 illustrates the VTRS-wide FOV that was contrasted with a smaller FOV available in the SH-60B OFT.
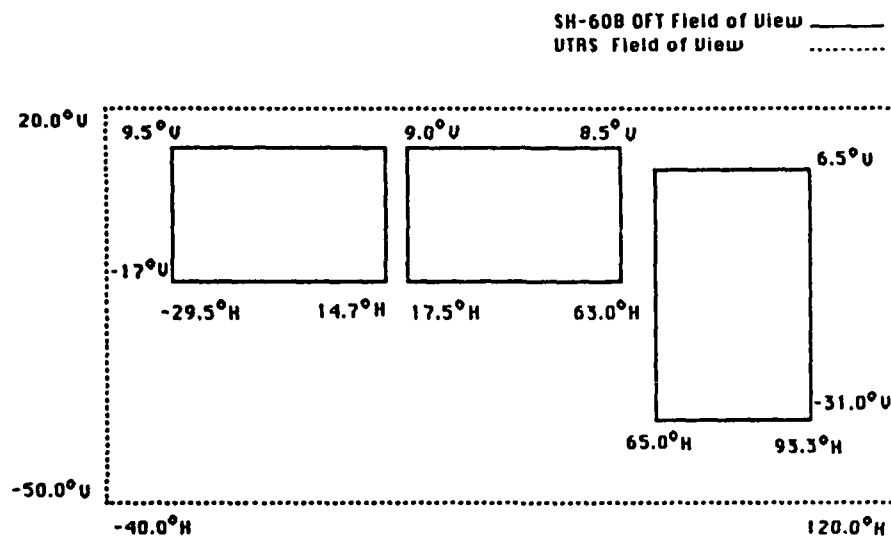
Figure 1. OFT and VTRS experimental fields of view.

Visual Transport Delay. Westra and Lintern (1985) contrasted visual transport delays of 117 and 217 msec, and obtained results indicating a performance advantage for the shorter lag condition in the hover, descent, and touchdown segments of the helicopter shipboard landing task. The current study also investigated the effects of visual transport delay and contrasted 117 msec versus 183 msec.

Dynamic Inflow. The SH-60B simulation model originally implemented at VTRS is typical in that induced velocity changes and subsequent rotor thrust changes respond "in toto" the instant rotor state changes. In reality, time is required for the flow field around each blade to change. The concept of dynamic inflow was introduced to the SH-60B rotor model to provide the transient behavior of rotor induced velocity and the subsequent behavior of rotor thrust. This feature makes the simulator more representative of the flight characteristics of the actual helicopter and was tested against the typical rotor model installed in operational trainers.

Procedure

Twelve experienced naval SH-60B helicopter pilots performed a precision hover (60 seconds over the deck of an FFG-7 frigate), and landing task in the VTOL simulator. Pilots received familiarization trials under several configurations before beginning experimental trials in order to acquaint them with the unique aspects of the flight simulator and to protect against initial learning effects. Each pilot completed 32

precision hover maneuvers over three days in sessions of eight trials each. Vertical wind gust disturbances (counterbalanced combinations of up and down) were introduced during each trial. The experiment was organized as a repeated-measures design with each pilot executing one-half of a full-factorial. The combination of two levels of five factors in the experiment resulted in 32 unique combinations or experimental conditions. Each pilot performed two trials on each of 16 experimental conditions, with each pair of pilots completing a full factorial. Thus, the full factorial design was replicated six times across pilots and twice within pilots.

Results

Results are presented here in condensed form. A complete presentation of results can be found in Westra, Sheppard, Jones, and Hettinger (in press). The analysis of variance tables shown here summarize the main effects of the experimental factors and also present mean differences (high minus low fidelity) between factor levels. Table 1 lists the factor levels in terms of "high" and "low" fidelity that were used to calculate mean differences.

Results for the precision hover task are presented in Tables 2 and 3. Lateral and longitudinal errors in the hover were defined as deviations from the center of the rapid securing device (RSD) on the landing pad. Vertical error was defined as deviations from a point 12 feet above the deck. Root-Mean-Square (RMS) error scores for lateral, longitudinal and vertical control in the hover are presented in Table 2. Two composite percent time-on-tolerance (TOT) scores for the three control dimensions combined are presented in Table 3.

There were highly significant effects on all performance measures (Tables 2 and 3) due to the field-of-view manipulation. Hover performance, measured in terms of RMS error for the lateral, longitudinal, and vertical axes, and the composite TOT scores, was significantly better with the VTRS FOV than with the OFT FOV. In fact, FOV accounted for almost 20% of the total variance in the RMS error data for the longitudinal and lateral dimensions of the hover (Table 2) and 30% of the variance in the composite TOT measures (Table 3).

The visual delay manipulation resulted in marginally significant effects on longitudinal and vertical control in the hover. In both cases, performance was better with the shorter delay of 117 msec (Table 2) Overall, the effects for visual lag were small but consistently favored the shorter lag (117 msec). The dynamic seat cueing factor had a marginally significant effect on longitudinal control in the hover (Table 2) and in composite TOT performance in the larger tolerance band (Table 3). In both cases however, performance was poorer when g-seat cueing was available. There were no performance differences attributed to the scene detail manipulation.

Dynamic inflow did not affect vertical control in the hover (which it was designed to affect), although it did significantly affect longitudinal control with better control observed under the standard rotor model condition (Table 2). An analysis of reaction times in response to the gust disturbances indicated a highly significant effect for dynamic inflow. On the average, pilots responded to the gust disturbances 97.3 msec faster when dynamic inflow was available compared to when it was not. No other significant equipment factor effects were observed for this measure.

Discussion

There were large field of view effects (up to 30% of the variance) on hover performance on all measures, with better performance occuring under the VTRS FOV.

TABLE 1. EXPERIMENTAL FACTOR LEVELS IN TERMS OF FIDELITY

Factor Levels	LOW	HIGH
Scene Detail	OFT scene	Upgraded VTRS scene
Field of View	OFT	VTRS
Dynamic Seat Cueing	Off	On
Dynamic Inflow	Standard Rotor Model	Updated Aero Model
Visual Delay	183 msec	117 msec

TABLE 2. ANALYSIS-OF-VARIANCE SUMMARIES FOR THE PRECISION HOVER TASK: RMS ERROR SCORES DURING HOVER

Source of Variance	df	RMS Error R/R[1] Longitudinal	Lateral	RMS Error Vertical
Dynamic Inflow	1	$0.50^2(1.4)$**	-0.12(0.3)	0.07(0.2)
Dynamic Seat Cueing	1	0.39(0.8)*	0.10(0.3)	0.07(0.2)
Scene Detail	1	-0.31(0.6)	-0.04(0.1)	-0.03(0.1)
Field of View	1	-2.08(19.4)**	-0.86(19.1)**	-0.40(4.6)*
Visual Delay	1	-0.38(0.7)*	0.02(0.0)	-0.15(0.7)*
Residual	297	(34.9)	(33.9)	(40.7)

TABLE 3. ANALYSIS-OF-VARIANCE SUMMARIES FOR THE PRECISION HOVER TASK: COMPOSITE TIME-ON-TOLERANCE SCORES DURING HOVER

Source of Variance	df	Composite X+Y+Z[3]	Composite X+Y+Z[4]
Dynamic Inflow	1	$-3.16^2(0.4)$	-0.41(0.0)
Dynamic Seat Cueing	1	-6.02(1.3)**	-1.43(0.3)
Scene Detail	1	2.54(0.2)	-0.26(0.0)
Field of View	1	32.83(33.5)**	17.82(29.2)**
Visual Delay	1	1.66(0.1)	2.29(0.5)
Residual	297	(35.3)	(38.9)

[1] R/R = Relative to the Rapid Securing Device
[2] Mean difference, i.e., mean for high level condition minus low level condition. (Percent-Variance-Accounted for in parentheses)
[3] X (longitudinal) = $\pm$ 5 ft; Y (lateral) = $\pm$ 3 ft; Z (vertical) = $\pm$ 3 ft
[4] X (longitudinal)= + 5 ft - 3 ft; Y (lateral) = $\pm$ 2 ft; Z (vertical) = $\pm$ 2 ft

*p < .05   **p < .01

Therefore, the VTRS FOV is recommended over the present OFT FOV. However, further research should consider an intermediate FOV (representing a FOV smaller than the VTRS FOV but larger than the OFT FOV) as the next step in determining the most cost-effective configuration for a training simulator. Dynamic seat cueing, on the other hand, showed no positive effect on hover performance. In fact, the g-seat appeared to have a detrimental effect on some hover performance measures. Thus, dynamic seat cueing as evaluated in this experiment is not recommended.

Visual transport delay had some small but consistent effects on hover performance, with better performance under the shorter display lag. In contrast to the effects found in Westra and Lintern's (1985) helicopter shipboard landing experiment, which compared 217 msec to 117 msec, it appears that 217 msec delay is more disruptive of performance than 183 msec. A delay of 183 msec, while not necessarily desirable, appears 'flyable' with only minor effects on performance compared to 117 msec. Although a display lag of 183 msec is marginally acceptable from a performance viewpoint, pilots rated this delay as inadequate for training novice pilots. In addition, data from a field transfer experiment (Westra et al., 1986) indicate positive transfer from the simulator to an operational task with a delay of 117 msec. We conclude that a display lag of 183 msec is acceptable from a performance viewpoint, but a shorter delay of 117 msec is recommended for training novice pilots.

Time series analyses revealed faster response time to wind gusts with the updated aero model. It did not appear to enhance vertical control performance as measured by RMS and TOT scores, and actually appeared to degrade performance slightly in the longitutinal dimension. However, the updated aero model did significantly improve performance in a helicopter shipboard landing task at VTRS (Westra et al., in press). Further research is suggested to define how modeling of rotor transient responses affects pilot hover performance. Scene detail did not have a significant effect on hover performance. Thus, the current deck and hanger wall markings in the OFT scene are considered adequate from a hover performance viewpoint.

References

Berbaum, K. S., & Kennedy, R. S. (1985). _An analysis of visual tasks in helicopter shipboard landing_ (NAVTRAEQUIPCEN 81-C-0105-19). Orlando, FL: Naval Training Systems Center.

McMillan, G. R., Martin, E. A., Flach, J. M., & Riccio, G. E. (1985, November). Advanced dynamic seats: An alternative to platform motion? _Proceedings of the 7th Interservice/Industry Training Equipment Conference_, Orlando, Fl.

Westra, D. P., & Lintern, G. (1985, September). _Simulator design features for helicopter landing on small ships. I. A performance study_ (NAVTRAEQUIPCEN 81-C-0105-13). Orlando, FL: Naval Training Equipment Center.

Westra, D. P., Lintern, G., Sheppard, D. J., Thomley, K. E., Mauk, R., Wightman, D. C., & Chambers, W. S. (1986). _Simulator design and instructional features for carrier landing: A field transfer study_ (NAVTRASYSCEN 85-C-0044-22). Orlando, FL: Naval Training Systems Center.

Westra, D. P., Sheppard, D. S., Jones, S. A., and Hettinger, L. J. (in press). _Simulator design features for helicopter shipboard landings. II: Performance experiments_ (NAVTRASYSCEN 85-D-0026-9). Orlando, Fl: Naval Training Systems Center.

An Assessment of an Integral Display and a Separable
Display: Which is the More Effective Sign ?

Mark S. Shurtleff, M.A
Dennis B. Beringer, Ph.D
New Mexico State University

Abstract

A comparison of three displays, which differed in their
degree of integrality, was conducted. The subjects performed a
rule-based task similar to what monitors of complex systems
routinely perform. In this task subjects identified a display
configuration (sign) as one that was previously learned or as
one that was novel. The integral display (polygon) was found to
be significantly better than the separable display (histogram)
as measured by the percentage of display configurations
identified correctly.

Increasingly humans are moving "out of the loop" in complex systems and
taking on the role of a monitor instead of a controller (Rasmussen, & Rouse,
1981). Highly automated systems require human intervention only when the
system goes out of tolerance due to automation failure. Thus, automation is
requiring the human to deal less and less with manual control and more and
more with cognitive interventions when automation fails.

Rasmussen (1983) has distinguished three categories of human performance
which occur in complex systems. The first, skill-based behavior, involves
sensory motor acts which, once initiated take place without conscious
control. The second, rule-based behavior, involves procedures stored in
memory that are derived from personal or vicarious experiences such as
training or word of mouth. Performance is goal orientated, but it is
structured by a stored rule. The third, knowledge-based behavior, involves
plans for action that are based on the operator's mental model of the system.
This type of behavior occurs when a situation is encountered for which no
known rules are available. Information displays are processed in a different
way depending upon the type of behavior the monitor is performing. Skill-
based behavior uses information as signals, e.g., as direct physical data.
Rule-based behavior uses information as signs, e.g., as learned patterns.
Knowledge-based behavior uses information as symbols, e.g., as concepts tied
to functional reasoning.

The task of the human monitor in complex systems, such as process control
plants and nuclear power plants, is to keep the system operating within
acceptable limits. The monitor, as such, is usually operating under rule-
based behavior and is using informational displays as signs. When the system
strays beyond tolerance limits the monitor must be able to identify the
configuration as one viewed previously during system operation or training.
If the configuration cannot be classified (is novel) then the monitor must
resort to knowledge-based behavior and employ the informational display as a
symbol. Thus, informational displays to be used in monitoring tasks must be

able to support the use of the display as both a sign and a symbol. Given adequate training, however, the display will be used most often as a sign.

Unfortunately, many studies evaluating displays for use in complex systems have not adequately addressed the full complexity of the monitoring task. Subjects, typically, were required to identify whether a certain configuration was in or out of tolerance (Munson & Horst, 1986; Carswell & Wickens, 1987). Such binary decisions, while comprising a part of the monitoring task, fail to capture its full complexity. It is necessary to know that a failure has occurred, but this is not sufficient. The monitor must also identify the type of failure that has occurred.

Complex systems possess many variables some of which are highly correlated. The information from a myriad of variables must be evaluated and their relationships assessed in order to identify the type of failure that has occurred. Integral displays are hypothesized to assist the monitor in identifying the type of system failure by configuring into a more salient property such as a unique shape (Pachella & Somers, 1978). This allows the relationships among variables to be perceived directly. Indeed, Carswell & Wickens (1987) have found an integral display to be superior to a separable display in the identification of system failure.

However, integral displays do have their drawbacks. Coury, et al., (1986) found that integral displays were not superior to a numeric display if different system states exhibited symptomatic behavior that resulted in a similar appearance (shape) of the integral display. Naveh-Benjamin & Pachella (1982) found that information contained in an integral display that was not relevant to the identification of a failed system could not be ignored. Furthermore, even when subjects were informed which information was irrelevant their perceptions of the display were still adversely affected.

At the other end of the continuum from integral displays are separable displays. These displays do not configure into a higher property that the subject can immediately process (Pachella & Somers, 1978). For example, a group of histograms may be thought of as a separable display. The height of each bar is processed separately. The bars do not configure into a unique pattern or shape as strongly as the polygon display does.

The goal of this study was to compare three graphical displays (polygon, constellation, histogram) which differed in their degree of integrality (see Figure 1) in a task which required subjects to identify the type of display configuration as one which was previously learned or as unknown. It was hypothesized that the integral display (polygon) would be superior to the separable display (histogram) as measured by the percentage of displays that were correctly identified. Furthermore, for the training task, in which subjects learned the names of the display configurations, it was expected that the integral display would be superior to the separable display as measured by the number of trials to criterion performance.

The second hypothesis is supported by the findings of Jacob, Egeth, & Bevan, (1976). They found subjects learned the names of two types of integral displays (Chernoff face display, polygon display) faster than they learned two types of separable displays (glyph display, numerical display).

346

However, they did not assess how well subjects could identify the learned configurations when intermixed with novel configurations. The polygon may be more easily recognized than the histogram display, but it may also be more subject to confusions with shapes that are reflections or otherwise similar to the learned shapes. With the histogram display, however, subjects would not be relying on global processing as they would be with the integral display (Munson & Horst, 1986) and, thus they may be less subject to confusions. The constellation display was created by displaying the endpoints of the polgon display without any connecting lines. This display is hypothesized to possess intermediate integrality, thus its performance was expected to be intermediate between the polygon display and the histogram display.
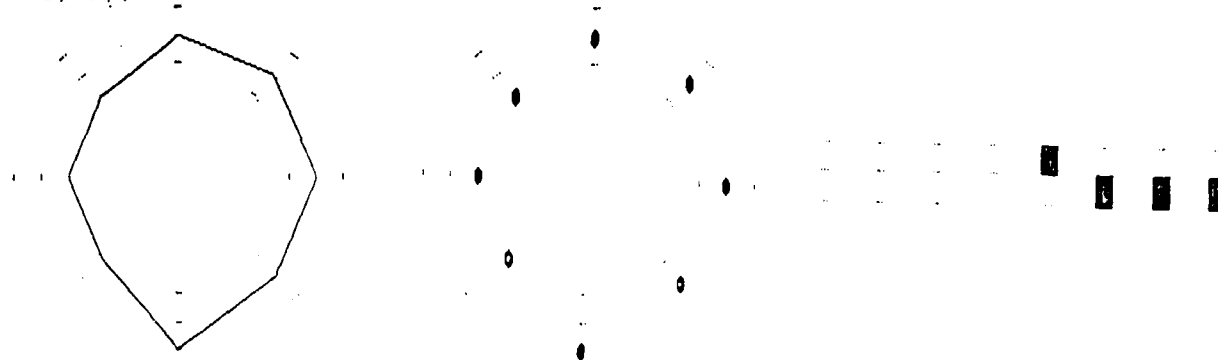
Figure 1. An example of the 'Alpha' display configuration for each type of display (left to right; polygon, constellation, histogram).

Method

Subjects/Apparatus

Twelve graduate students from the psychology department at New Mexico State University volunteered to participate in the experiment. All twelve subjects tested met the criterion for task one and completed task two. The apparatus used in this study was an IDS sixteen-bit digital microcomputer equipped with a monochrome monitor measuring 28 cm diagonally. The computer was programmed in PASCAL to present the stimuli and to record the dependent measure. Subjects indicated their responses by depressing a key on the computer's numeric keypad.

Procedure

Two tasks were employed in the present experiment. In the first task subjects learned the names of five unique display configurations. In the second task subjects identified the five learned display configurations which were randomly mixed with five novel display configurations.

Subjects were randomly assigned to one of the display types (polygon, constellation, histogram). The names (alpha, beta, gamma, delta, epsilon) and the variable values were identical for each type of display. Each display had eight variable indicators. In turn, each of the eight indicators could take on one of three levels. The display configurations were selected nonsystematically by consulting a random number table to obtain a value (1, 2, or 3) for each variable indicator.

347

Subjects were first instructed that they were to learn the names of five display configurations. The subjects viewed a printed copy of the five named configurations for 30 seconds. Afterwards they responded to stimuli which were presented individually on the computer screen. Subjects viewed a fixation cross for 2 seconds, then the display was shown. Subjects were given as much time as they desired to respond, then they were given knowledge of results (KR) for 5 seconds. If they were correct the KR was the word 'Right'. However, if they were incorrect the KR was the word 'Wrong' along with the display's correct name. Criterion performance was defined as naming the five shapes correctly twice in a row.

Upon attainment of the criterion the subjects then performed the second task. In this task subjects were required to identify the five previously learned configurations intermixed with five novel display configurations. Subjects were instructed to identify a known display configuration by responding with its name and to identify a novel display configuration by responding 'Other'. A total of 40 display configurations were presented. The subjects first viewed a fixation cross for 2 seconds. Then the display was shown for as long as the subject took to identify it. Once identified, the subjects received KR, displayed for five seconds. The word 'Right' was displayed for correct responses while the word 'Wrong' was displayed for incorrect responses.

Results

The dependent measure of number of trials to criterion in the learning task was analyzed with a one-way ANOVA. Also the percent correct identifications in the identification task was analyzed with a another one-way ANOVA.

For the learning task no significant effects were found. Although the raw data does show a trend favoring the polygon display (mean trials to criterion: polygon, 6.5; constellation, 8.5; histogram, 9.0), there was too much variability in the data to reach significance.

For the second task the effect of display type was found to be $F(2,9) = 3.85$, $p < .06$. Given the small sample size, this result is indicative of a significant display effect. Thus three pairwise, nonorthogonal LSD tests were performed to find if the differences between the means were significant. A significant difference was found between the polygon and the histogram display ($p < .05$) (see Figure 2).
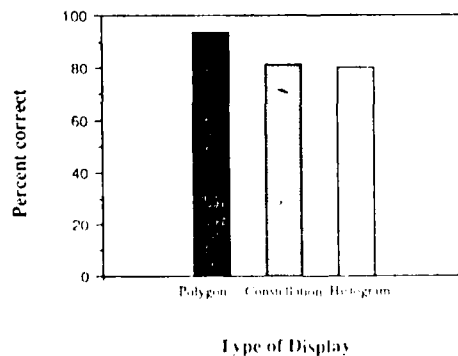
Figure 2. Mean percent correct identifications by display type.

348

Discussion

The results of this experiment tend to support the advantage of an integral display (polygon) in the identification task. The accuracy of the subjects using the polygon display was impressive -- all four subjects performed at 88% correct or above. Clearly, this performance was quite different than the subjects who used the histogram display where the best performance was 85% correct. The separable display most probably was encoded in its constituent parts making the classification task difficult as compared with the polygon display. The superiority of polygon display over a histogram display has also been demonstrated by Carswell, & Wickens (1987) in a dynamic system with a task that required information integration (identification of a system failure).

This study adds to the literature that supports the advantages of an integral display in a monitoring task in which the display was being processed as a sign. Another important aspect of the monitoring task, which this study did not assess, involves knowledge-based performance in which a novel failure configuration is encountered and must be classified so that corrective action can be taken. In this case the display would have to be processed as a symbol. There is a need to assess integral displays in this situation. Integral displays appear to be effective signs, but further research is required to determine whether or not they are effective symbols.

References

Carswell, C. M., & Wickens, C. D. (1987). Information integration and the object display: An interaction of task demands and display superiority. Ergonomics, 30(3), 511-527.

Coury, B. G., Boulotte, M. D., Zubritzky, M. C., & Fisher, D. L. (1986). Classification of multidimensional data: Evaluating displays to aid identification of system state. In N. Morris (Ed.), Proceedings of the 1986 I.E.E.E International Conference on Systems, Man, and Cybernetics, I.E.E.E., 1503-1508.

Jacob, R. K., Egeth, H. E., & Bevan, W. (1976). The face as a data display, Human Factors, 18(2), 189-199.

Munson, R. C., & Horst, R. L., (1986). Evidence for global processing of complex visual displays. Proceedings of the Human Factors Society, 30th Annual Meeting, Human Factors Society, 776-780.

Naveh-Benjamin, M., & Pachella, R. G. (1982). The effect of complexity on interpreting "Chernoff" faces, Human Factors, 24(1), 11-18.

Pachella, R. G., & Somers, P. (1978). The development of integrated multidimensional displays, Naval Research Review, 31(11), 19-32.

Rasmussen, J. (1983). Skills, rules, and knowledge; Signals, signs, and symbols, and other distinctions in human performance models, I.E.E.E Transactions on Systems, Man, and Cybernetics, 13(3), 257-266.

Rasmussen, J., & Rouse, W. B. (1981). (Eds.). Human detection and diagnosis of system failures. New York: Plenum Press.

International Symbol Set Development

Marilyn Vala-Rossi, Ronald D. Lyons, and Currell L. Pattie
Human Factors Engineering/System Safety
General Dynamics Land Systems

Abstract

This study developed and validated a methodology for the
development of symbols to enhance or replace existing text labels
for controls and functions in military tracked vehicles.
Forty-five (45) General Dynamics engineers were asked for sketches
of twenty-three (23) controls and functions identified from M60,
M1, and Tank Test Bed (TTB). Three symbols for each control and
function were derived from the characteristics occuring most
frequently in the sketches. To arrive at one symbol for each
control and function these symbols were then tested for response
times and errors on five tasks: Association Task, Rating Task,
Paired-Associate Learning Task, Reaction Time Task, and Secondary
Task. The results from twelve (12) General Dynamics engineers and
eighteen (18) soldiers from Ft Knox produced a final set of
symbols that possessed consistant characteristics. They were
simple in design, pictorial, and very specific to that particular
control or function.

Use of equipment manufactured by foreign countries, increases in
information load, limited space on displays, and requirements for rapid
responses all place high communication demands on label and indicators. The
communication of information through symbols has a number of advantages over
written, textual information and can help meet this challenge. One obvious
advantage is that symbols typically require considerably less space to
communicate information. This is particularly important in displays and
computer displays. In addition researchers have found that symbols are
preceived more rapidly (Janda and Volk, 1934; Carr, McCauley, Sperber, and
Parmelee, 1982) and accurately (Walker, Nicolay, and Stearns, 1965) than their
textual counterparts. Symbols have been found to be superior to text in
conditions where the information is visually degraded (Ells and Dewar, 1979),
viewed at a distance (Dewar and Ells, 1974), or when distractions from other
tasks occur (King, 1975). A properly conceived set of symbols potentially
will reduce response time, minimize erors, increase learning transfer, and
decrease training time.

Some methods of deriving symbols are better than others (Mudd and Karsh,
1961; Karsh and Mudd, 1962; Green, 1979). The most productive method for
generating symbols is to derive the symbols empirically. This is referred to
as the "population-stereotype production" technique (Mudd and Karsh, 1961;
Howell and Fuchs, 1968; Green, 1979). This technique requires a group of
subjects to draw simple pictures to represent a word; for example, "fuel
pump". Drawings are collected and tallied with respect to certain
"characteristics". The most frequently occuring characteristics then are
combined to form one symbol. In studies that have directly compared
empirically derived symbols to symbols using other developmental techniques,
such as graphic arts, the empirical derivation proved superior in conveying
symbol meaning (Mudd and Karsh, 1961; Karsh and Mudd, 1962, Green, 1979). It
is this method which was used to derive the twenty-three (23) symbols which
were tested in this study. An example of these symbols is shown in Figure 1.

CONTROLS AND FUNCTIONS   A   B   C

Ammunition Select

Autoloader Stop/Continue

Bilge Pump
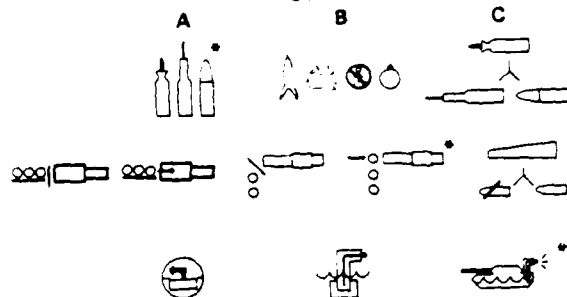
Figure 1. Symbol Samples

Method

Subjects

The subjects were twelve (12) employees from General Dynamics Land
Systems Division with previous tracked vehicle experience either as tank
commanders, drivers, gunners, platoon leaders, or mechanics. Eighteen (18) M1
and Bradley tank commanders and drivers from Ft Knox were also used in the
study.

Procedure

Subjects completed five tasks; Association Task, Rating Task,
Paired-Associate Learning Task, Reaction Time Task, and Secondary Task.

Association Task: The degree to which the symbols were associated to the
message they represented was examined in the following manner: the sixty-nine
test symbols were randomly ordered on a page. The experimenter read a
scenario and the subject was asked to give the number or numbers of the
symbols which described the scenario. They were not told how many symbols
there were for each control and function, and their responses were not
limited. There were twenty-three (23) different scenarios.

Rating Task: A rating task was also used to gather magnitude estimates of
how well the symbols conveyed the meaning of the control or function they
represented (communicativeness). The symbols were randomly assigned to a
14x21 inch board so viewing would be easy for the subject. The subject was
also given a numbered sheet of control and function messages that corresponded
to the symbols on the board. The subject was asked to look at the symbol,
then look at the control or function message it represented, and rate its
communicativeness on a scale of one (low communicativeness) to five (high
communicativeness). Then, they were asked to rate their confidence in that
rating, also on a scale of one to five. Subjects were given an unlimited
amount of time to make evaluations.

Paired-Associate Learning Task: A paired-associate learning task was given
to collect estimates on how quickly the symbols could be learned. In this
task the subject was required to pair a symbol shown on a CRT screen with the
correct name of the control or function it represented. If the subject said
the right answer, the subject was told he was correct and went on to the next
symbol. If the subject said the wrong answer he was told the correct answer
and was asked to concentrate on remembering that correct label. The task
proceeded in this manner until the subject had correctly named the symbol
three times. That symbol was then removed from the display sequence. This
continued until the criteria had been met for all the symbols.

Reaction Time Task: Using the groupings generated in the paired-associate
learning task, three random samples were generated for the three control and
function groups. Each sample consisted of the three symbols for a particular

control or function displayed five times, for a total of fifteen correct responses. Fifteen symbols were then randomly drawn from the other two sets and used as follows. The subject sat approximately 24 inches from a CRT screen. The "A" and "D" keys on the key board were labeled SAME and DIFF. These positions were counterbalanced between subjects so half the subjects used SAME on the "A" and "DIFF" on the "D", and the other half used the opposite arrangement. After pressing the return key, a text label for one of the controls or functions would appear on the screen, for example "BLACKOUT LIGHTS". The subject was given a short verbal definition and asked to remember the label. He was then told to press the space bar and a symbol would appear. Using the keys, he would then respond as quickly as possible whether the symbol was the same or different from the label. Subjects were told that occasionally they would make errors, but that was necessary to assure they were answering as quickly as possible.

Secondary Task: Although the reaction time task provided baseline reaction times for the symbols, it was of interest to determine what happened to the reaction time to symbols when variables inherent to the operation of a tank were introduced as independent variables. For example, during normal operation of a tracked vehicle, the operator must monitor and respond to numerous information displays. This demands a great deal of his attention. He also experiences fatigue from his constant alert, observing operational condition. Such circumstances could alter the crewman's responses to the symbols.

To evaluate these variables, an attention demanding task was created that required the subject to track a tank around the four quadrants of the screen at one second intervals. The tracking keys were "Y" which indicated the tank was in the upper left corner, "I" (upper right corner), "B" (lower left corner), and "M" (lower right corner). The subject was told that accuracy in tracking the tank was important and to try to be as accurate as possible. They also were told that part of the task would be similar to the reaction time task they performed previously; i.e., they would see a label for one of the twenty-three (23) controls and functions, and would be required to make a same or different decision using the same keys that were used in the reaction time task. As in the reaction time task, decision speed. The tracking task resumed immediately after they pressed the SAME or DIFF key.

The same random orders used in the reaction time task were used in the secondary task; however, these were counterbalanced within subjects. The order of presentation changed although they saw the same controls and functions as the previous task.

Informal Interview: After completing all the tasks, subjects were asked to answer questions designed to provide information on any confusion they may have experienced when viewing the symbols. Subjects also were asked to provide feedback on improving the symbols.

<center>Results</center>

While the data gathered from each individual task was useful, a method was needed to employ this information to choose one final symbol from the three possible symbol candidates. The following approach was used to arrive at a final symbol for each control and function: Three symbol candidates for each control and fucntion were rank ordered based on performance within each task. The mean ranking for each symbol was calculated and is shown in Table 4. The highest mean ranked candidate for each control and function was chosen as the final symbol. When ties occurred, the symbol that received the highest mean rating on the rating task was selected.

Discussion

The use of this technique consistently discriminated between good symbols and poor symbols. The results show that symbols receiving high scores on the rating task also were associated with lower errors and faster response times. The final symbol set possessed consistent characteristics. The symbols that performed the best were simple in design, pictorial, and very specific to that particular control or function. For example, a "good" symbol for smoke generator would be a simple line drawing of a tank with smoke coming out of the rear. The results show that symbols that were abstract in design, received poorer overall scores, as well as poorer scores in the paired-associate learning task. The implications for military applications are apparent using the paired-associative learning task as an indicator of training difficulty. Training of a symbolic language would require the user to learn a new language system. This would place additional burden on the training process.

The symbols that were included in this study that were not developed from population stereotypes also showed overall poor performance. These symbols are currently in use on an armored tracked personnel carrier (APC). They were designed by a graphics artist and never tested empirically. While most are pictorial in content, the messages they were designed to transmit are unclear, particularly for users unfamiliar with the weapon system. Such a symbol set would increase confusion and errors, rather than minimize them.

Symbols that possessed generic characteristics (i.e. a flame to show fire or smoke) also performed poorly. These symbols were responsible for the highest number of association task errors, indicating they were confused with other controls and functions, thereby defeating the goal of minimizing errors. Also, test subjects did not find these symbols particularly informative, and rated them lower than the other two symbol candidates. This may be a result of the symbol overgeneralizing, and not identifying the particular control or function.

Future testing of symbols utilizing this technique must take place to assure this technique can be applied to a variety of symbol types and populations. The final symbol set found in this study should be tested in foreign speaking populations to determine if similar results occur. A different symbol set should be generated and tested using the same technique to see if it discriminates between the symbols as it did in this study.

From the information obtained in this study, it can be concluded that some symbol types are responded to better than others. Symbols must transmit information efficiently to meet the goals of reduced response time, minimized errors, increased learning transfer, and decreased training time. This study demonstrated that techniques similar to those described in this report yield symbols which meet these criteria.

References

Carr, T.H., McCauley, C., Sperber, R.D., Parmelee, C.M. Words, pictures, and priming; On semantic activation, conscious identification, and the automaticity of information processing. Journal of Experimental Psychology:Human Perception and Performance. 1982, 8, 757-775.

Dewar, R.E. and Ells, J.G. A comparison of three methods for evaluating traffic signs. Transportation Research Record. 1974, 503, 38-47.

Dreyfuss, H. Symbol Sourcebook, New York City, N.Y. McGraw-Hill, 1972.

Ells, J.G. and Dewar, R.E. Rapid comprehension of verbal and symbolic traffic sign messages. Human Factors, 1979, 21, 161-168.

Green, P. and Pew, R.W. Evaluating pictorial symbols: An automotive application. Human Factors, 1978, 20, 103-114.

Green, P. Development of pictographic symbols for vehicle controls and displays. Society of Automotive Engineers, Paper Number 790383, 1979.

Howell, W.C. and Fuchs, A.H. Population stereotypy in code design. Organizational Behavior and Human Performance, 1968, 3, 310-339.

Janda, H.F. and Volk, W.N. Effectiveness of various highway signs. National Academy of Sciences, National Research Council, Highway Research Board Proceedings, 1934, 14, 442-447.

Karsh, R. and Mudd, S.A. Design of a picture language to identify vehicle controls. III. A comparative evaluation of selected picture-symbol designs. Aberdeen Proving Ground, Maryland: U.S. Army Human Engineering Laboratory Technical Memorandum, 15-62, August, 1962.

King, L.E. and George, Z.H. A further investigation of symbol versus word highway signs. a paper presented at the Annual Meeting of the Human Factors Society, New York City, October, 1971.

King, L.E. Recognition of symbol and word traffic signs. Journal of Safety Research, 1975, 7, 80-84.

Utilization of Color Contrast in Electronic Tactical Display Maps

Ronald D. Lyons, Jeffrey M. Powers, & Marilyn Vala-Rossi
Human Factors Engineering/System Safety
General Dynamics Land Systems
Dean G. Purcell & Frank Schieber
Department of Psychology
Oakland University

Abstract

The present study investigated the effects of target color on
the ability of subjects to identify targets against color borders.
To determine those effects target presentation was made under two
primary background conditions; solid colors and target placement
at the intersection of two colors. Ten male volunteers were asked
to search for two types of target stimuli, a solid target or line
target. The subjects were required to push a lever to indicate
which side of a divided computer screen the target appeared. The
results indicate that subjects performed best with black targets
while light-green yielded the worst performance. The results also
showed solid targets yielded better performance than line targets
on borders.

The literature on the effects of color upon visual search is well
documented (Christ, 1975). What is less well know are the requirements
concerning which target and background hues to use for color coded
computer-generated terrain maps (see Spiker, Rogers, & Cicinelli; 1986).
Spiker et al. (1986) used a visual search task with stimuli superimposed on
a computer generated topographic map. Their study demonstrated advantages
for certain target-background color combinations. Spiker et al. (1986)
failed to test the effects of target placement on the border of two
different colors. This condition is likely to occur in a real world
situation. The present study investigated the effects of target placement
on color borders. In order to determine such effects, target presentation
was made under two primary background conditions. One condition consisted
of a solid color and the other consisted of target placement at the
intersection of two colors. In addition, the effects of target color,
target type (Line vs Solid), and background color were investigated.

Method

Subjects

Ten males served as Subjects in this experiment. All participants were
screened to assure optimal acuity and color vision. Subjects were paid for
their participation.

Apparatus

Stimuli were presented via a NEC high-resolution color monitor driven by
a PC clone. A Matrox Model PIP-1024 graphics frame-buffer was used to
generate the stimuli.

Stimuli

Two general types of target and distractor stimuli were used; solid and line-drawn symbols. An example of each is shown in Figure 1. The solid symbols were created for the experiment. The line-drawn symbols were designed to closely approximate those symbols presented in the NATO Standardization Agreement.

Solid Triangle   Anti-Tank Symbol

Figure 1. Example: Solid Target and Line Target

The target stimulus colors consisted of Black, White, Magenta, Cyan, Orange, Blue-Green, Light Green and Red. These eight colors were selected to sample across the visible spectrum. The short wavelengths were under-sampled because of the known limitations of the visual system in those wavelengths.

The target stimuli were presented against ten background conditions. Two general types of backgrounds were used: solid color, and two colors arranged in a series of diamonds (i.e., a diagonal checkerboard pattern). For the two-color, patterned backgrounds, targets and their distractors were presented at the intersection (border) of the two colors. Patterned backgrounds consisted of Blue-on-Green, Brown-on-Green, Yellow-on-Green, Brown-on-Blue, Yellow-on-Blue and Yellow-on-Brown, while solid colored backgrounds were either Green, Blue, Brown or Yellow. They were chosen primarily to represent existing map color coding schemes.

Targets and distractors were randomly positioned from trial to trial with the constraint that targets must appear at the intersection or border of the two-color backgrounds. Targets and distracters were presented at a luminance of 25 cd/m2 (nominal).

Procedure

The subject's task was to search for two types of target stimuli: a solid target or a NATO-line target. One of these two targets was presented on each trial. Seven distractors were presented along with the target. The distractor stimuli changed from trial to trial.

The factorial combination of the two target types, eight target colors, and ten backgrounds created 160 separate conditions. Each condition was presented 12 times for a total of 1920 trials per subject. Subjects received 45 minutes of practice. At the end of that time a check was made to determine if a subject's average reaction time had reached asymptote.

During stimulus presentation the screen was bisected by a vertical black line. At the start of each trial a central fixation cross appeared as a signal to the subject to initiate the trial. Coincident with the fixation cross was the background condition for that trial. When the start button was pressed, the target, distractors, and background stimuli were presented and remained "on" until the subject responded. The subject responded by pressing a lever to the right or left to indicate the position of the target stimulus relative to the black line. All trials were self paced. In all experiments, error trials were replaced by rerunning them at the end of the

session. Reaction time analyses were based upon correct responses only.

Results

For the primary analysis, a repeated measures ANOVA (see Table 1) was performed on the reaction time data. While no main effect was found for target type the three way interaction of target type, target color, and background was significant (p<0.01). This interaction will be the focus of discussion.

OVERALL ANOVA

SOURCE	DEGREES OF FREEDOM	MEAN SQUARE	F	PROB.
A	1	258140	1.10	0.3214
	9	234489		
B	7	547909	31.64	p<0.01
	63	17310		
AB	7	32458	2.41	0.0302
	63	13495		
C	9	804824	48.24	p<0.01
	81	16684		
AC	9	80245	8.74	p<0.01
	81	9184		
BC	63	179467	13.19	p<0.01
	567	13602		
ABC	63	49595	4.67	p<0.01
	567	10626		

A=Target type
B=Target color
C=Background

Table 1. Overall ANOVA

A criteria was established using the target color data to examine which target colors were related to the fastest response times. Since subjects demonstrated the fastest response times with both the solid and line-drawn black targets, the slowest black target response time (800 msec) was used as the criteria. The matrix in Figure 2 indicates which color targets meet this criteria.
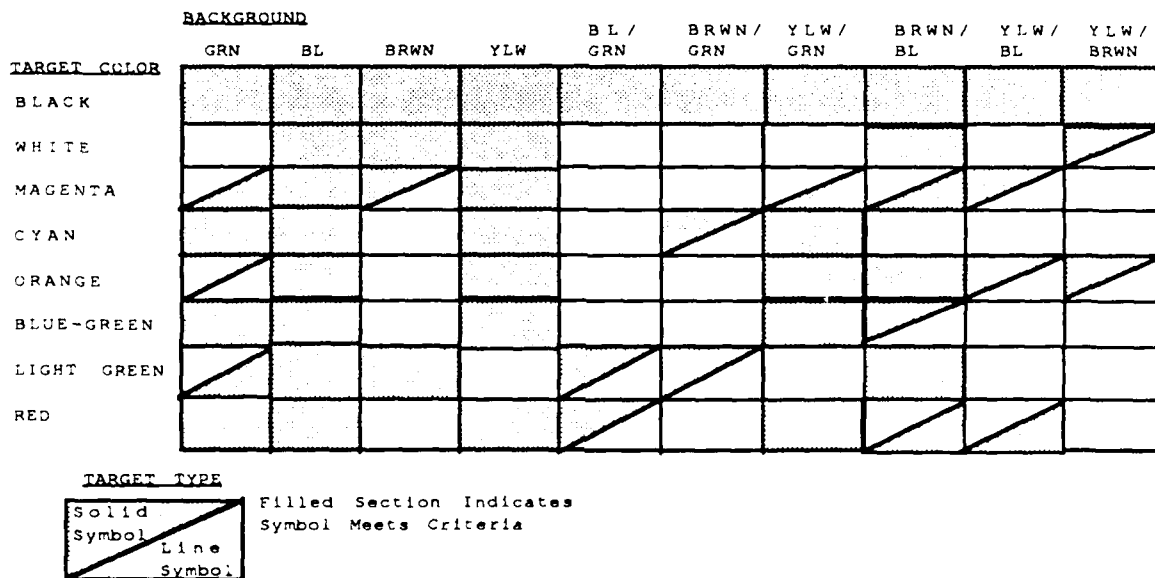
Figure 2. Target Color and Type vs. Background

Discussion

Based on the above results a number of recommendations can be made for using color on an electronic tactical map. The significant three way interaction indicates that all three of these variables must be taken into consideration when designing for color displays. However, the background colors representing different terrain features (i.e., blue for water, green for land) will remain constant for tactical maps. Thus the color and type of symbols overlaid on these backgrounds should be of most concern.

The matrix in Figure 2 indicates which target colors and types were associated with the fastest response times. As the matrix shows black is the best target color for both solid and lined symbols on both the solid and border backgrounds. Magenta and Orange targets were also associated with the fast response times in both solid and border background conditions. Fewer colors met the criteria on the borders than on the solid backgrounds. For example, white, blue-green, and cyan only met the criteria on two borders. These colors should be avoided for critical information.

Target type was considered less critical than color for overall performance. Seventy-five percent (75%) of the solid and line targets on the solid backgrounds met the criteria. The ten percent (10%) advantage for solid targets on border backgrounds vs. line targets (45% vs.35%) indicates solid targets would be recommended.

Line symbol performance may be improved by incorporating black, (recall that black was associated with the best target performance), into the color symbols. For example, a blue rectangle with an oval inside symbolizes a friendly armor unit. This blue line symbol could be filled in with black pixels, thus enhancing the symbol but not altering the blue color. This could theoretically cause colored symbols to perform at levels equal to black symbols. A second experiment was run to test this theory.

Experiment 2: In Experiment 1 it was generally found that Solid symbols were associated with better performance than line symbols. Experiment 2 was designed to determine if it would be possible to increase performance levels for the lined symbols used in Experiment 1. Spiker et al. (1986) report that symbol identification is enhanced by bordering the symbols with black. Rather than border the symbols with black the symbols were filled with black. Such symbols have an advantage of isolating the symbol from its background while covering up less of the background than bordering it with black. Experiment 2 was essentially the same as Experiment 1 with the exception that only Black-Filled symbols were used. The independent variables in Experiment 2 were Target Color and Background. These variables were manipulated at the same level as described for Experiment 1.

Method

Subjects

Eight male volunteers served as subjects. All participants were screened to assure optimal visual acuity and color vision.

Apparatus

The apparartus was the same as used in Experiment 1.

Stimuli

The stimulus symbols were the same as the line-drawn symbols used in Experiment 1 with the exception that the interstices of these symbols were filled with black. The background colors used were also the same as Experiment 1.

Procedure

The procedure for Experiment 2 was similar to Experiment 1. In Experiment 2 the factorial combination of eight target colors, and ten backgrounds created eighty (80) separate conditions. Each condition was represented twelve (12) times for a total of nine hundred-sixty (960) trials per subject. Distractors either were the same type as the targets (solid or line), opposite the target type, or a mixture of the two. The experiment was run the same as Experiment 1.

Results

For the primary analysis, a repeated measures ANOVA (see Table 2) found the two-way interaction between foreground and background to be significant. Figure 3 shows that the black-filled line symbols were associated with overall faster response times than the average color of the other symbol types (solid and line).

OVERALL ANOVA

SOURCE	DEGREES OF FREEDOM	MEAN SQUARE	F	PROB.
A	7	53466	3.16	0.0077
	49	16898		
B	9	114249	10.10	p<0.01
	63	11311		
AB	63	15244	2.64	p<0.01
	441	5777		

A=Foreground
B=Background

Table 2. Overall ANOVA
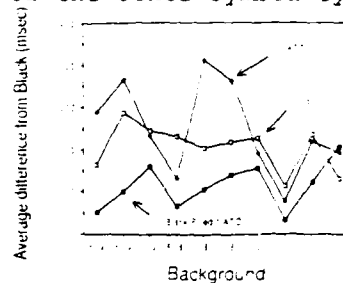
Figure 3. Black-Filled vs. Other Symbols

Discussion

Experiment 2 demonstrates the possiblity of improving response time performance of colored, line symbols by filling them with black pixels. This is beneficial to design of color electronic tactical maps for two reasons. First, it provides flexiblity in the symbol types that can be used on the maps. The symbols currently used by the military are the NATO line symbols. Unless the military changes their symbol system, these are the symbols that would be required for electronic tactical maps. The ability to enhance these symbols on an electronic system would ease the transition to a new system by providing improved information already familiar to him. Secondly, color would no longer be as critical, since black would be incorporated into every symbol, thus providing colored symbols the benefits of black targets. Use of these techniques when designing electronic tactical maps will improve the chances of providing a better system for the user.

References
Christ, R. E. (1975). Review and analysis of color coding research for visual displays. Human Factors, 17, 542-570.

Spiker, A., Rogers, S. P., & Cicinelli, J. (1986). Selecting colour codes for a computer-generated topographic map based on perception experiments and functional requirements. Ergonomics, 29, 1313-1328.

Anthropometric Design Criteria
Correction Variables for Military Aircraft

Harv Gregoire, Ph.D.
Rannell Dahl, M.S.
University of Southern California

Abstract

Significant errors in standards of anthropometric design criteria for tactical military aircraft cockpits have been identified. These errors result from insufficient consideration for previously undefined variables relative to posture, restraint system items, and aviators personal protection equipment (McConville and Laubach, 1978).

For years, specific standards of anthropometric criteria have been applied in the design of aircrew stations in the research, development, and acquisition of military aircraft (McConville and Lauback, 1978). Three major problems in the application of such criteria are addressed in the present paper; namely, (a) the optimum design eye position in a cockpit, (b) design features that influence the reachability of cockpit controls, and (c) effects that the required items of aviators' equipment have on the design of cockpit geometry.

Specifically, the current research quantifies differences existing among (a) those human dimensions traditionally obtained by standard anthropometric techniques (Kroemer, 1978), (b) the displacement in posture that results from items of personal equipment (McCormick and Sanders, 1982), and (c) the postures actually assumed by individuals attired for flight and further influenced by the actual aircraft ejection seat and restraint system (Gregoire 1975; McCormick, 1982).

Due to advances in technology, the identification of an optimum design eye position in cockpits is becoming more important than it has been. Visual information that was previously displayed at various places throughout the cockpit is increasingly displayed on Heads Up Displays (HUD's) (Walter and Hoover, 1976). Although HUD's provide an increased capability for outside surveillance, these displays as well as other advanced state-of-the-art electro-optical devices require more than ever that the most accurate definition of optimum design eye position be achieved for tactical military aircraft of the future.

Reach constraints will also become more critical in cockpit design due to the higher performance capabilities of future combat aircraft (Chambers 1963). The higher "g" environments of future combat aircraft will demand a significantly smaller margin of error in control placement than before.

Method

Subjects

Of the research population sampled, 53 were males and 51 were females.

Females were included as a result of recent changes in policy which includes training and utilization of women as pilots in military aviation.

Apparatus

A special facility was designed and constructed to research the aircrew station criterion problem. The facility featured (a) an anthropometric measuring chair typical of the type used to measure aviators' dimensions such as total sitting height, sitting shoulder height, fingertip reach, etc., (b) an ejection seat which included a restraint system, a parachute, survival kit elements, and flight control stick, all from a tactical military jet aircraft and all arranged to simulate the actual man-seat to control-stick relationships found in a typical military tactical jet aircraft, and (c) a specially designed visual grid to take measurements. The anthropometric chair, ejection seat, and measuring grids were aligned vertically and horizontally on the common axis of the Seat Reference Point (SRP), the intersection of the seat surface and back surface tangent lines (Huchingson, 1981). Additionally, equipment items and flight clothing were provided to enable subjects to be realistically attired for in-flight operations.

Procedure

The first measurement phase consisted of measuring and recording anthropometric data as has historically been done to attain standard anthropometric design criteria. In the next phase, the subject was attired in a standard torso harness, flying helmet, oxygen mask, floatation device, and flying gloves. The subject was then instructed to sit in the anthropometric chair and assume as similar a posture as possible to that of the first measurement phase. Those measurements directly affected by the wearing of the aviators equipment were then retaken. Differences were then computed between the original dimensions and those altered by the, aviators equipment in order to determine the amount of volume added or posture displaced by each item of equipment.

In the third phase of measurement, the subject, completely attired for flight, was seated in the ejection seat and strapped in securely. The subject was then instructed to simulate maintaining the aircraft in straight and level flight by holding the control stick centered with the right hand while reaching about the cockpit with the left hand for specified reach targets. The purpose for requiring simulated flight was twofold. First, mechanical advantage from pushing or pulling against the ejection seat structure was prevented. Secondly, validity was enhanced in that during flight, aircraft control via physical manipulation of the control stick is normally required to maintain or change aircraft attitude while other subsystem contol tasks must be performed. Next, a reach target apparatus was aligned in front of the subject. The target locations were selected as being representative of normal and emergency control locations in a typical tactical military aircraft.

The last phase of measurement was then initiated with the measurement of total sitting height, sitting eye height, sitting shoulder height, and buttock-to-knee length. The reach envelope was then evaluated using the reach target apparatus. With the shoulder harness locked, the subject's reach dimensions were recorded in tow modes. The first mode was normal non-stretching reach; the second mode was maximum stretch. In the maximum stretch mode, the subject exerted as much as possible against the shoulder straps and torso harness in order to reach as far as possible with a locked shoulder harness. Differences

were then computed between the normal reach and maximum stretch modes in order
to quantify how much farther a restrained individual could reach as a function
of target location, torso harness stretch, and other restraint system para-
meters.

Results

In the total sample population, the inflight eye position was 3.94 cm
below and 2.54 cm forward of the traditionally determined eye position. The
major engineering design implications include a relocation of seat adjust-
ment locus about a redefined seat reference point or modification of seat
and canopy design configurations which will permit approximately 3.80 cm
higher seating relative to present criteria.

Normal straight ahead non-stretching reach was 3.31 cm less from an
ejection seat than from an anthropometric chair aligned with the same seat
reference point. Maximum stretch for the restraint system (the Douglas
Escapac Ejection Seat used in A-4 and A-7 type tactical jet aircraft) used in
this research was, on the average, 5.46 cm more than the measurements from the
anthropometric chair. It should be emphasized that maximum stretch is
difficult if not impossible to achieve at certain critical times during high
"g" flight operations such as catapult launch from an aircraft carrier, air
combat maneuvering, weapons delivery pull-out, etc. (Damon, Stoudt, McFarland,
1966).

Differences between normal and maximum stretch to various cockpit areas
relative to left shoulder location were 5.26 cm for 15 degrees to the right,
and 5.11 cm for 15 degrees to the left. The engineering design implications
are that the reach capability of an aviator is rigidly and quite uniformly
restricted by typical ejection-seat torso-restraint system. It is imperative
that normal and emergency controls be placed well within the reach envelope
of the smallest members of the intended user population. In the population
sampled, the straight-ahead fingertip reach differences between normal and
maximum stretch reach for males and females were similar, being 8.96 cm for
males and 8.51 cm for females. However, the difference between the average
maximum reaches of the males and females sampled was large, being 6.07 cm.

Discussion

The design implications resulting from this study of aircrew design
criteria are significant. It appears that anthropometric criteria previously
specified in some military standards has not been realistic. The designation
of the design eye position and seat-to-control distances as well as the
methods of measuring these parameters have not always included the variables
of posture slouch, torso restraint systems, and posture displacement resulting
from aviators equipment. These variables significantly alter the position of
morphological features in actual flight environments.
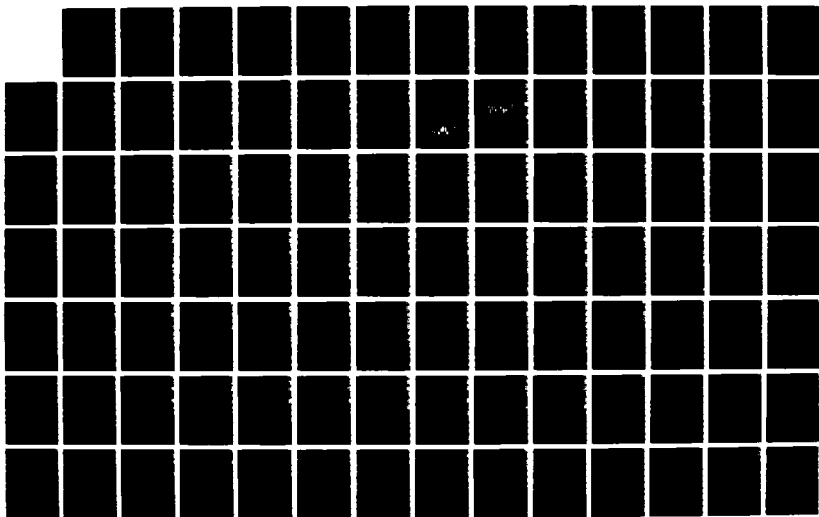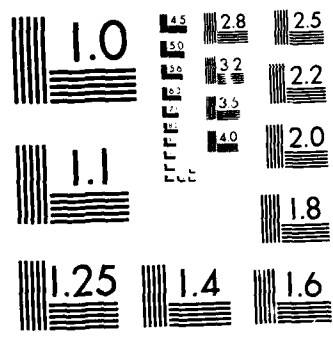
References

Chambers, R.M. (1963). Operator Performance in Acceleration Environments. In Burns, R. M., Chambers and E. Handler (eds.). Unusual Environments and Human Behavior. New York: The Free Press, 193-320.

Damon, A., Stoudt, H. W., and McFarland, R. A. (1966). The Human Body in Equipment Design. Cambridge, MA: Harvard.

Gregoire, H. (1975). Doctoral Dissertation, Lubbock, TX: Texas Tech. University.

Huchingson, R. D. (1981). New Horizons for Human Factors in Design. New York: McGraw-Hill, p.201.

Kroemer, K. H. E. (1978). Functional Anthropometry. Proceedings of the Human Factors Society, pp. 680-683.

McConville, J. T. and L. L. Laubach (1978). Anthropometry. Anthropometric Source Book, NASA Reference Publication 1024, NASA Scientific and Technical Office, 1.

McCormick, E. J. and Sanders, M. S. (1982). Human Factors in Engineering and Design. New York: McGraw-Hill, p. 324 and p. 288.

Walter, J. F., Jr., and G. W. Hoover (1976). The Maneuvering Flight Path Display. Proceedings of the Third Advanced Aircrew Display Symposium, Maryland: Patuxent River Naval Air Test Center.

Successful Integration of Human Engineering Principles
in System Design: A Discussion and Recommendations

1LT Jay A. Horn
Air Force Institute of Technology
Wright State University

James M. Barnaba
Aeronautical Systems Division
Wright-Patterson Air Force Base

ABSTRACT

The need for human factors engineering in the
acquisition process is receiving increased attention but
is still misunderstood by some. This paper examines
several facets of human factors engineering that deserve
further discussion. The reasons for implementing a human
factors program and utilizing human factors resources are
identified. Five requirements necessary for the successful
integration of an effective human factors program in
today's complex weapons system acquisition environment are
then identified and discussed. Finally, recommendations of
what must be done to increase awareness and application of
human factors in the procurement cycle are offered.

INTRODUCTION

The past has seen human factors (HF) considerations disregarded and
even excluded from the design of new weapons systems. This paper presents
working guidelines for the successful integration of human factors
efforts into the system acquisition program. Areas that will be
addressed include the personnel involved, the program manager's
orientation towards human factors, the contractual requirements levied on
the contractor, and the activities and interactions necessary to insure
successful project completion. The discussions and recommendations
presented here are intended to benefit government program management
personnel through increased awareness of the HF field, but they are also
relevant to HF engineers and contractors.

There are several compelling reasons why persons who direct the
acquisition of modern weapons systems must be cognizant of the role of
human factors in system development. The Government Accounting Office
(GAO) released a report describing the increased need for HF, logistics
support, and quality assurance in the systems acquisition process (1).
Air Force Regulation (AFR) 800-15 directs and explains the use of HF in
the acquisition process. The new specialty code identifier 2675A
shred-out, for behavioral scientists specializing in HF, was recently
created by the Air Force, reflecting the need to identify personnel with
human factors engineering experience.

The Air Force Systems Command (AFSC) has named HF as a special

interest item for the Inspector General in 1987. General Bernard P. Randolph, AFSC Commander, in his Commander's Policies letter, has identified meeting the user's needs and requirements as one of the top priority goals for the Command. This increased emphasis on HF by the GAO, the Department of Defense, and Air Force leadership clearly demonstrates that top-level management is committed to effective implementation of human factors considerations into the procurement process.

Increased cognitive and workload demands placed upon the weapon system operator and maintainer due to the system's increased capabilities in information display and control present persistent and difficult challenges (2). The solutions to these challenges can only be found with the application of human factors principles during system design. The GAO report cited above named three facts supporting the inclusion of HF in system design:

1. Human errors account for at least 50% of the failures of major weapons systems.
2. Neglect of human factors in early planning has caused serious and costly problems.
3. There has been considerable effort to adapt man to the constraints built into hardware instead of using human factors design criteria to initially design a system compatible with the human's limitations.

There are several aspects of the current system design process that result in the problems noted above. First among these is the implementation of immature technology simply because it can be done without regard for how the technology enhances (or degrades) the system. Air Force Manual 1-1, Basic Aerospace Doctrine of the USAF, describes this problem of balancing the rewards of tommorrow's technology versus today's need for realistic war-fighting capability (3). Human factors engineers can provide the program manager with information vital to determining the trade-offs between these two concerns.

Over-reliance on the human operator's ability to adapt and inattention to the intended user's desires and requirements are also significant sources of problems. Ignorance of the intended user population's skill levels and attributes, as well as postponement or complete disregard for HF considerations to the detriment of the system, are also causes of the problems revealed by the GAO study. A competent human factors engineer can provide the information necessary to remedy these problems.

With these aspects in mind, it becomes obvious that human factors programs must be implemented in today's technologically advanced weapons system procurement arena. Effective use of HF during the design and test stages of the procurement will result in long term savings of time, money, and lives, as well as increased operational performance and user satisfaction.

Indeed, the emphasis placed upon HF will differ with each program. It is the responsibility of the human factors engineer (HFE) to provide program management with the information necessary to determine the needs

of the particular program. Armed with this information, the program
manager can successfully employ his HF resources for the benefit of both
the system and the Air Force.

DISCUSSION

A strong and effective human factors program includes these assets:

1. Government Expertise. The government human factors expert must
be trained and experienced in both HF principles and the DOD procurement
process.

2. Contractor Expertise. The contractor must provide a
knowledgeable and experienced team of HF professionals who are acquainted
with the government acquisition process.

3. Program Management Support. The government and contractor
program managers must both realize the need and importance of human
factors in the development of a successful system, and have confidence in
and provide the necessary support to their HF experts.

4. Complete Contractual Documentation. A thorough and well-written
system specification stating the requirements of the program and a
statement of work (SOW) that clearly addresses the HF program and
includes appropriate human factors contract data requirements list (CDRL)
items.

5. Government-Contractor Communication. Frequent exchange of
information and ideas must occur between all cognizant personnel to
monitor program progress and respond to development questions.

The government HF expert needs to be brought on board as soon as the
program manager begins developing the request for proposal (RFP) for a
new system. The program manager must include an HF expert early to
ensure that HF concerns are adequately addressed in the RFP. Contact can
be made with the user in order to determine his needs and allow the the
HF engineer to fully understand the task and mission requirements of the
system. This understanding can be developed through on-site visits to
the users under the auspices of the Air Force's Blue Two program or
through similiar operational visiting programs. The Blue Two program
allows USAF and contractor personnel to travel to the end-user's
facilities and get first-hand information of how the item will be used
and maintained. At this time, the HF expert will be able to determine
what the HF requirements will be and advise the program manager of the
amount of HF support necessary for the program.

The HF expert will then be able to task the contractor through the
SOW with appropriate CDRL items to ensure that the user's requirements
are met. A well-written RFP and SOW with appropriate CDRL items will
result in better proposals from the contractors, and allow the strengths
and weaknesses of each human factors program to be identified.
Documented government requirements for CDRLs and specific project
milestone completion dates allow the contractor human factors specialist
to do his job better since he can now approach his program manager with
options to meet the documented requirements. A contractor that has an

experienced human factors engineering staff and is familiar with the federal acquisition process will tend to reduce the risk associated with the program.

The HF expert who developed the RFP and source selection criteria should be used in evaluating the contractors' proposals in order to reduce the risk of accepting a proposal that does not meet the system's requirements. The HF evaluator must provide sufficient rationale to explain each offeror's human factors program, in terms of meeting the source selection criteria, and the potential risks associated with each.

Desirable proposals will clearly outline a human factors program with scheduled milestones including (but not limited to) design reviews, technical interchange meetings, Blue Two visits, user's group conferences, and system tests. The schedule should also identify any human factors tasks being performed before the first major design review. The proposal should also describe the relationships between the HF group and other engineering disciplines to insure incorporation of HF concerns throughout the design process. Up-front analysis and inclusion of human factors concerns into the system design reduce risk and allow for timely completion of the system development phase.

Upon contract award, the government and contractor HF experts must meet to discuss the philosophy of the program and begin planning for the most immediate concerns of the program. It is important at this time to build a good working relationship to facilitate the flow of information between both parties. A regular dialogue must be maintained between the government and contractor to insure that problems are rapidly addressed and solved. This also allows the government HF expert to know the status of the system's development while allowing the contractor a ready source of information about the user's needs and the system's requirements. Regular communication between government and contractor HF experts will lead to a more effective HF program, and consequently, a better operational system.

RECOMMENDATIONS

To realize the benefits gained by implementation of an effective human factors program, there are several actions the procuring agency must pursue:

1. The procuring agency must hire sufficient numbers of well trained human factors engineers and provide opportunities for them to maintain currency in their field and pursue higher education.

2. The procuring agency must provide training and opportunities for human factors experts to become well-versed in the federal acquisition process, techniques of writing the system specification, SOW, and RFP, and source selection evaluation.

3. Program managers must be made aware of the need for human factors expertise in the development of all weapons systems and must make use of the human factors expertise available to them. AFSC Regulation 550-10 directs program managers to ponder "each decision in light of the

user's needs (4)." Consulting their human factors resources will give the program managers insight into the user's perspective.

  4. The government must insure that the contractor realizes the need for trained human factors expertise on the engineering team.

  Although the government does hire trained personnel to become HF experts and provides training opportunities, there is room for improvement. Current manning appears to be adequate for today's demand. But, as increasing emphasis is placed upon human factors, more and better trained personnel will be needed. More time and money must be available to pursue additional education to allow the HF engineers to maintain their skills and keep abreast of the latest techniques. Several training courses are offered to instruct new hires about the systems acquisition cycle, but although every effort is made to send everyone, the quality of each course differs, and work commitments sometimes preclude attendance.

  It is important to overcome the stereotypical view (held by both managers and engineers alike) that human factors is just "common sense" or gold-plating. Human factors engineers do make meaningful and vital contributions to every program's success. We cannot afford to ignore the evidence any longer. As the impact of command initiatives is felt and as the level of technology continues to rise, more DOD personnel must come to realize the importance of adequately addressing human factors issues in system development or the loss of life, equipment, and operational capability will continue. In today's battlefield of compressed time and space, we must possess the most capability we can get for our procurement dollars.

REFERENCES

1. Effectiveness of U.S. Forces Can Be Increased Through Improved Weapon System Design. Government Accounting Office Report PSAD-81-17. January 29, 1981.

2. Too Much, Too Soon: Information Overload. Article in IEEE Spectrum, Volume 24, Number 6, pp. 51-55. June, 1987.

3. Air Force Manual 1-1, Basic Aerospace Doctrine of the United States Air Force.

4. Air force Systems Command Regulation 550-10. Dated 24 July 1987.

THE ROLE OF HUMAN FACTORS IN
SYSTEMS ACQUISITION

Major James A. Boyless, PE
USAF Academy

Abstract

Modernizing combat forces by acquiring new weapon systems to meet the need for improved operational capabilities is one of the most important, challenging, and complex tasks faced by the U.S. Government (Rich and Dews, 1986). To increase the probability of achieving these improved operational capabilities, the inclusion of Human Factors Engineering (HFE) in the entire acquisition interval is essential. This paper introduces the acquisition interval, the directives and regulations mandating HFE, and the HFE products and activities in each stage of the interval.

Modernizing combat forces by acquiring new weapon systems inherently demands increasing complexity to meet those operational capabilities expected in future defense forces. Since men/women will be operating, maintaining and training with these modern systems, a closer scrutiny of the complex relationship between man/woman and the system must start early in the acquisition interval.

The interval, consisting of the mission area analysis, concept exploration, demonstration and validation, full scale development, and production, must have human factors engineers as an integral part of the "program management team." In fact, DoD Directive 5000.1 states that DoD acquisition of major defense systems must be carried out efficiently and effectively to achieve the operational objectives of the U.S. Armed Forces in their support of national policies and objectives,...(DoD Directive 5000.1, 1982). These operational objectives are further delineated into operational effectiveness and suitability with their respective definitions.

Operational effectiveness of the system is defined as the overall degree of mission accomplishment used by representative personnel...and environment in the planned operational employment of the system (DoDD 5000.1, 1982). The directive continues with the definition of suitability as the degree which a system can be placed satisfactorly in field use, with consideration to availability, compatability, transportability, interoperability, reliability, wartime usage rates, maintainability, safety, <u>human</u>

factors(underlined for emphasis), manpower supportability, logistic supportability and training requirements (DoDD 5000.1, 1982).

With the overriding concern of acquiring weapon systems which meet operational objectives, this paper will discuss the role of human factors and present the HFE products and activities necessary in the acquisition interval.

Discussion

The role of human factors in the acquisition interval is multi-faceted. As members of the program management team, human factors engineeris are an integral part of the research and development planning, conceptual study efforts, exploratory, advanced, and engineering development projects,...where the intended end product has human performance as an integral part(AFR 800-15, 1974).

The objective of this role, stated in the MIL-STD 721B (1966) definition of human engineering, a subtask of human factors, is that area of human factors which applies scientific knowledge to the design of items to achieve effective man-machine integration and utilization.

Scientific knowledge in human factors is the accumulated data that includes the capabilities and limitations of humans in conjunction with the proposed system and/or equipment design. Data from diverse yet related disciplines such as Psychology, Physiology, Industrial Engineering, Computer Science, Mathematics, and Anthropology enable human factors engineers to approach system objectives from an empirical base.

These system objectives include enhancing the effectiveness and efficiency with which work and other activities are carried out, as well as, enhancing certain desirable human values such as improved safety, reduced fatigue and stress, increased comfort and job satisfaction (Sanders and McCormick, 1987).

The application of human factors occurs, as previously mentioned, in all stages of the acquisition interval. In each stage, human factors methodology results in principal HFE products. These are instruments that detail the work in future stages. Figure 1 shows each stage and the HFE products that are further delineated into HFE activities (See Figure 2). Since

371

these activities are to numerous to fully discuss in this paper
readers are urged to consult both Geer and Sawyer.
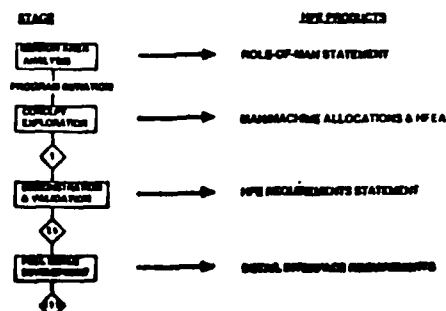
Figure 1. Stages and HFE Products (Boyless, 1986)

MISSION AREA ANALYSIS

HFE ACTIVITIES

- IDENTIFY SYSTEM FUNCTIONS, OPTIONS, REQUIRED CAPABILITIES

- PROJECT OPERATIONAL/ENVIRONMENTAL CONDITIONS

- DETERMINE ADVANTAGES/DISADVANTAGES OF VARIOUS OPERATOR/MAINTAINER ROLE ALTERNATIVES

- CONSIDER TENTATIVE MAN-MACHINE ALLOCATIONS (FUNCTION ANALYSIS)

DEMONSTRATION & VALIDATION

HFE ACTIVITIES

- CONDUCT TASK ANALYSIS TO DETERMINE:
  - INFORMATION AND CONTROL REQUIREMENTS
  - OPERATOR/MAINTAINER SKILL REQUIREMENTS
  - OPERATIONAL SEQUENCES
  - WORK FLOW, WORKLOAD (TIME-LINE ANALYSIS)
  - WORK STATION/SPACE ARRANGEMENTS
  - CONSOLE DESIGN
  - CONTROLS/DISPLAYS
  - INGRESS/EGRESS
  - MAINTENANCE

- USE ABOVE TO WRITE HFE REQUIREMENTS

CONCEPT EXPLORATION

HFE ACTIVITIES

- LIST FUNCTIONAL CRITERIA

- MATCH HUMAN CAPABILITIES AGAINST CRITERIA

- LIST MOST FEASIBLE MAN-MACHINE ALLOCATIONS FOR EACH FUNCTION

- MATCH FIND ALLOCATIONS AGAINST TRADEOFF CRITERIA

- FUNCTION ALLOCATION

FULL SCALE DEVELOPMENT

HFE ACTIVITIES

- MONITOR USE OF HFE REQUIREMENTS IN PROTOTYPE

- EVALUATE SYSTEM DESIGN UNDER VARIED OPERATIONAL AND ENVIRONMENTAL CONDITIONS

- FINE TUNE TO ACHIEVE DETAILED MAN-MACHINE INTERFACE DESIGN

Figure 2. Stage Activities.

Although the activities shown in figure 2 are to numerous to
fully describe in this this paper, example activities will be

discussed to give the reader an idea of what occurs during a particular stage.

The first stage of the acquisition interval is the Mission Area Analysis. In this stage, there are several activities that identify, project, determine, and consider system functions, requirements, operational/environmental conditions, operator/maintainer roles, and man-machine allocations. Geer(1981) presents a word picture of the analysis process that starts with the system mission as described by a baseline scenario. The mission objective and functions that must be performed by the system are identified, described, and sequenced. These functions are then analyzed to determine their proper allocation to personnel, software, or equipment.

In the Concept Exploration Stage, the activities that occur include listing functional criteria, matching human capabilities against that criteria, listing the most feasible man-machine allocations for each function, match/find allocations against tradeoff criteria, then allocate. Once allocated, the personnel functions are further analyzed to determine the specific operator/maintainer tasks or roles which must be performed to accomplish the functions (Geer,1981).

The Demonstration and Validation Stage follows to demonstrate the selected system concept and test its feasibility with regard to mission requirements. In this stage, there are two basic tasks: the task analysis and writing the HFE requirements (Sawyer, 1981).

The final stage, Full Scale Development, is the stage most important to the human factors engineer. It is here that the human factors engineer must refine and evaluate the system concept for fulfillment of HFE requirements, conformity to HFE design criteria, quantify system performance, and detect poor design or procedures (Sawyer, 1981).

Summary

DoD Directives and Air Force Regulations are used to direct personnel in acquiring modern weapon systems to meet operational capabilities. These capabilities further guide personnel to acquire systems that are effective, meet mission need, compatible with human performance capabilities and limitations yet are safe,

reliable, maintainable, and suitable.

Application of relevant human factors data is mandated in research and development, conceptual study efforts, exploratory, advanced, and engineering development. Acquiring weapon systems is a challenging, complex task and the program management team that includes human factors engineers will increase the probability of effective and operational systems.

In each phase of the acquisition interval there is a human factors product that is supported by HFE activities. By applying scientific knowledge from related disciplines, these activities will support weapon systems that meet mission needs.

References

Rich, M. and Dews, E. (1986). Improving the Military Acquisition
    Process, Lessons from Rand Research, R-3373-AF/RC,
    Santa Monica, CA.

DoD Directive 5000.1 (1982). "Major Systems Acquisition,"
    Department of Defense.

AFR 800-15 (1974). "Human Factors Engineering and Management,"
    Wash D.C., Department of Air Force.

MIL-STD-721B (1966). "Definitions of Effectiveness Terms for
    Reliability, Maintainability, Human Factors, and Safety,"
    Department of Defense.

Sanders, M.S. and McCormick, E. J.(1987). Human Factors in
    Engineering and Design, 6th Ed, New York: McGraw-Hill.

Boyless, J. A., Payne, D. and Kaplan, J. (1986). "Human Factors
    and System Acquisition: A Workshop," Human Factors Society
    Annual Mtg, Dayton, OH.

Geer, C. W. (1981). Human Engineering Guide, TR-81-35, AFAMRL,
    WPAFB, OH (NTIS No. A108 643).

Sawyer, C. R. et al (1981). Measuring and Enhancing the
    Contribution of Human Factors in Military System Development:
    Case Studies of the Application of Impact Assessment
    Methodologies, US Army Research Institute, Alexandria, VA.

The Effects of Competitive and Cooperative Reward Structures
on Working Memory

Jeffrey Todd Wimmer, C2C
Arthur Patrick Wunder, C2C
United States Air Force Academy

Abstract
This study sought to determine the effects
of competitive and cooperative reward structures
on working memory. We felt differing reward
structures would cause significant differences
in results on the Sternberg memory task. Our
data revealed no significant difference in
average reaction times and no difference in
percentage correct. However, the variance
between individual subjects reaction times
were significantly higher for the competitive
group than for the cooperative group.

American public education systems, universities, and
almost every type of teaching or training have relied,
almost exclusively, upon competitive reward structures.
That is, students compete against each other for extrinsic
rewards such as grades, promotions or simply recognition.
However, a large body of contemporary research and opinion
has suggested distinct advantages for cooperative reward
structures.

> Cooperation is the basic phenomenon
> that distinguishes our species. It's the
> underpinning for everything. That includes
> socialism and capitalism, business and family.
> Any society, even one obsessed with compe-
> tition, is predicated on people cooperating
> with one another.(Johnson & Johnson, 1987,p.53)

David and Roger Johnson, whose research on cooperation
includes more than eighty original studies, profess
"Children who learn cooperatively...compared to those who
learn competitively or independently...learn better."
(Johnson & Johnson, 1987, p.53) "In fact, the more complex
the learning task, the better cooperation fared."(Johnson &
Johnson, 1987, p.54) The Johnson brothers also suggest that
a cooperative reward srtucture would not hinder or hold back
gifted students. Gifted students would provide explanations
and actually help lead the group. This simple verbal review
reinforces knowledge and the act of providing explanations
is indicative of a deep level of processing and
assimilation. Students also appear to have much more fun
and be much more enthusiastic about teaching each other and
learning together.

We were fascinated by some of the implications of the
Johnsons' assertions and were interested if their might be

an effect due to varying reward structures at the micro level of human performance. We felt that assigning rewards by group performance would increase overall performance, since each individual would feel responsible not only to himself, but also to the group for his individual performance. In addition to the increase in average performance, there was reason to expect a difference in the variance in the two groups. Under competitive reward systems, distinctive roles become self-fulfilling profecies; talented people may become vindictively competitive while less talented people may give up. This effect would be shown by less variance in average reaction times in the cooperative group relative to the competitive group.

A computer version of Sternberg's working memory task was selected as the experimental tool to test these hypotheses.

Method

Subjects

Thirty-six cadets participated in this experiment. Three groups of six performed the task under a cooperative reward structure and three groups of six under competitive reward structure. All subjects were given the privilege of ordering pizza and missing a training meal in order to participate in this study.

Apparatus

The experiment was conducted in the Behavioral Sciences and Leadership Laboratory at the United States Air Force Academy. We used six Zenith Z-248 computers on a single table to administer the Sternberg task. The subjects faced each other and a "blip", loud enough for all subjects to hear, went off whenever an incorrect response was made. This was to facilitate the group's awareness of each others performance.

Procedure

The same proctor was used on all trials and all subjects started the task at seven o'clock AM during a normal school day. A group of six subjects was randomly selected from the thirty-six volunteers, ushered into the computer testing laboratory, and told to read the instructions on the screen pertaining to the Sternberg task. None of the subjects knew what the task was before entering the laboratory. After asking if there were any questions, the proctor told the subjects to go ahead with a practice trial in order to familiarize themselves with the task. When finished, the reaction times of the group practices were recorded and then one of the two reward structures was explained.

Competition

The competitive groups were told the individual with the quickest reaction time and at least 90% accuracy would receive a small reward. After the first trial, the subject with the quickest reaction time was given three candy bars, his time announced to the group, and the subjects proceeded with the next trial.

Cooperation

The cooperative groups' reaction times on the practice trials were averaged. These groups were told that if they achieved a reaction time 3% lower than their previous average (without anyone going below 90% accuracy), they all would receive a small reward. If the subjects achieved the target of 3% lower than the previous reaction time then each subject received a candy bar and the group was again asked to better their average reaction time by 3%. If the group failed to achieve the target reaction time or if any single subject went below 90% accuracy, then no one was rewarded and the group was allowed to try at the unachieved target reaction time again. In order to transition between target times of differing memory set sizes, the widely-accepted additive factor 35 milliseconds for each additional element of the memory set was used.

Results

Mean Reaction Time vs Memory Set Size

Memory Set Size

□ Cooperation   ◇ Competition
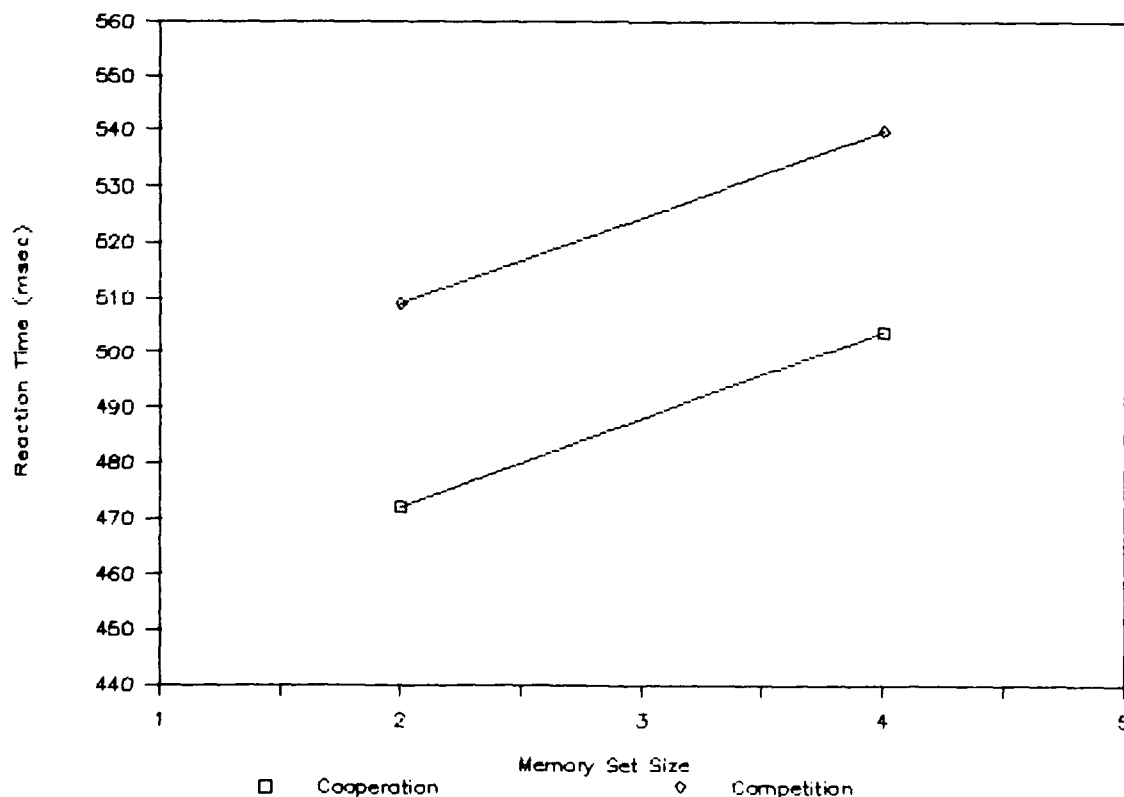
An examination of the raw data suggests that the effect of memory set size was independent of the effects of cooperation - vs - competition. This independence corroborates our method of simply adding a fixed time for the larger memory set.

We used the performance data from the practice trials in order to test the equality of our randomly assigned groups. Neither mean average reaction time ($t=1.59$, $df=34$, $p>.10$), nor variance ($F_{17,18}=1.45$, $p>.10$) differed significantly but the cooperative group was slightly quicker and less variant.

The average reaction times and standard deviations for the four criteria trials, after the differing reward structures had been presented were 454.0 msec (SD=48.4) for the cooperative group and 481.3 msec (SD=91.9) for the competitive group. This advantage for the cooperative group was not significant ($t=1.11$, $df=34$, $p>.10$).

The percentage correct was not influenced by the reward structures either. The competitive reward structure resulted in a mean of 92.7% accuracy and the cooperative reward structure resulted in a mean of 92.5% accuracy. Our second hypothesis, that a cooperative reward structure would reduce the variance between subjects was supported. Infact, the variance for subjects in the competitive condition was over three and a half times as great as for subjects in the cooperative condition ($F_{17,17} = 3.61$, $p<.01$).

Discussion

Several aspects of these findings are interesting. The clear lack of evidence for interactive effects of reward structure (competitive - vs - cooperative) and memory set size (2 or 4) suggests that these two influences are additive and thus independent. Memory set size is frequently associated with the central executive portion of working memory which in turn is most closely related to attention and "consciousness." (Baddley, 1983) This tenuous string of suppositions hints that some of the effects of competitive reward structures are not conscious. The performance advantage for the cooperative group was not statistically significant. While this result did not fully support our hypothesis, the difference was in the correct direction. The common belief that competition improves performance received no support. The one significant difference we found in the two reward structures involved the variance in individual performance within the two groups. The cooperative groups were considerably less variant.

This basic research has several pedagogical implications. If the negative effects of competition are at least partially subconscious or preconscious, verbal reports of students or faculty concerning the effects of reward structure may be less than valid and complete. Also,

attempts to attenuate the negative effects of competition by verbal arguments are likely to be somewhat ineffective.

The increase in variance in performance associated with competitive reward structures would seem to greatly complicate the educational process. To the extent education relies on communication and communication relies on shared experience; greater variance suggests lowered educational efficiency. Competition creates an enviroment where student's different experiences and levels of understanding make it almost impossible for the instructor to communicate to everyone with the same words and also makes it very difficult for students to communicate with each other even if they choose to. The importance of this effect is particularly important in light of the fact that "inherent individual variability", is one of the necessary assumptions underlying the widely accepted practice of "grading the curve". Our findings suggest that such "inherent variability" may infact be an artifact of the competitive reward structure itself.

References

Baddeley, A.D. (1983). "Working Memory" in Philosophical Transactions of the Royal Society of London. London.

Eysenck, M. (1984). A Handbook of Cognitive Psychology. London: Lawrence Erlbaum Associates Publishers.

Hinkle, D., Jurs, S. & Wiersma, W. (1979). Applied Statistics for the Behavioral Sciences. Boston: Houghton Mifflin Company.

Johnson, D. & Johnson, R. (1987). It's Hard to Get Left Out of a Pair. Psychology Today. October, 52-57.

Sternberg, S. (1969). The Discovery of Processing Stages: Extensions of Donders' Method. Acta Psychologica, 30, 276-315.

Cooperative Learning and Interdependent Grading in Physics Class at the United States Air Force Academy[1]

Lt Colonel Richard E. Swanson, Ph.D.
Department of Physics
United States Air Force Academy

Two groups of students in an introductory physics course were taught using cooperative learning techniques which emphasized student centered activities and peer interactions. Individuals were rewarded for group performance by using interdependent grading. One group performed as expected and showed performance and attitude improvement. The other group performed well but had a dramatic drop off at the end of the course which may be due to the formation of counterproductive cliques within the group.

The effects of competition and cooperation on education and training become critically important when considered in the military environment. In a recent evaluation, Wood and Flannery (1986) discuss competition at the Air Force Academy in three main mission areas: military leadership, academics, and athletics. They point out that competition is an important and necessary part of the Academy experience, yet its negative effects must be understood and dealt with. They encourage the use of cooperative techniques in all environments. Of particular interest to this study are their statements about academics. They state that "cadets view academics as primarily an individual effort, and therefore, an individual competition" (Wood and Flannery, 1986, p.12) and "teamwork or cooperative ... efforts are virtually non-existent in the academic arena." (Wood and Flannery, 1986, p.31) They discuss the negative aspects of curve grading and students' perception that individual good performance hurts the group by raising the mean.

These sentiments are echoed in the literature. A work cosponsored by the American Association for Higher Education and the Education Commission of the States (Chickering and Gamson, 1987) states that one of the primary principles of "good practice in undergraduate education" is that it "develops reciprocity and cooperation among students. "Bruffee (1987) discusses the need to allow students to form peer learning groups by altering classroom instruction to foster cooperation and Kohn (1986) discusses the negative effects of competition on the educational environment and stresses the need for cooperation. He states, "It is not simply the competitiveness of an individual that undermines achievement. A structure that demands competition tends to have the same effect ..." (Kohn, 1986, p.24) and points out that, "The vast difference between striving for excellence and striving for victory is clear in many fields" (Kohn, 1986, p.28).

The purpose of this study is to report on the effects of using

----

[1]Special thanks to Lt Col Dave Porter for his valuable contributions.

cooperative learning techniques in the introductory physics course, Physics 110, at the Air Force Academy. I tried to emphasize all aspects of cooperative learning groups as outlined by Johnson and Johnson (1986).

Method

Subjects

In the fall semester, 1987, Physics 110 consisted of 602 primarily sophomore students divided into 28 classes of about 21 students per class. Each class was taught by one of the 11 instructors in the course. I chose three of the classes for the study. Two of them I will call Group 1 and the other I will call Group 2. I was the instructor for both groups. Group 1 consisted of two classes of 22 students each. Their average grade point average (GPA) coming into the course did not differ significantly from the population average. Group 2 consisted of one class of 20 students. This group is treated separately from Group 1 since they were also part of a related program called the IDEA (Interdisciplinary Education at the Academy) program. Students in this group were also in the same introductory military history class. The history instructor and I worked together to encourage cross-discipline integration. One hope was that the increased interaction time among these students would allow them to form a more effective peer learning group.

Instruments

Student performance was evaluated using three major examinations (called Graded Reviews or GR's) and the final examination. All students in the population took these same examinations and the group averages were compared to the average scores for the entire course (Figure 1).

Student attitudes were evaluated using an end of the course survey developed by the Physics Department and the Department of Behavioral Sciences and Leadership. Again, group average responses were compared to the population average (Figure 2).

Student comments about their course and class were solicited. They responded to an electronic mail message sent to them in their dorm rooms. They were asked to send their instructor an electronic mail message telling: what they liked, what they didn't like, and what they would change.

Procedure

The classes were administered for Group 1 exactly like the rest of the course. Group 2 was treated differently only in that they also met for the same history class and in that sense were more aware that they were a special group. The students in both groups were encouraged to cooperate in primarily two ways: modifying in-class activities and using interdependent grading.

Much more class time was given to student centered tasks that were designed to promote positive interdependence as well a individual accountability (Johnson and Johnson, 1986). For example, rather than a lecture on motion, groups would be asked to measure and evaluate the motion of a diver

in a videodisc sequence. After a quiz was announced, students would be given
the opportunity to help each other prepare. A reasonable estimate is that
actual lecturing comprised about one third of the time that is typically given
to lecture in this course. Instructor interaction with individuals and
groups increased accordingly.

Interdependent grading is an administrative mechanism for rewarding the
individual for the improved performance of the group. Prior to each quiz or
examination, each group is given a target score that they are expected to
achieve. If the group performance on the exam exceeds that target, each
individual in the group has points added to his or her score. Alternatively,
if the group performance is less than the target, points are taken away from
each student. The student who performs well now helps everyone else in the
class rather than hurting them by "raising the mean". It is to everyone's
benefit that the poorer students are helped to perform better.

Results

The results on the three GR's and final examination are shown in Figure
1. The scores reported reflect actual performance in that any interdependent
grading points added or subtracted are not included. Both groups performed
close to or below the population average for the first two GR's. These
differences were not statistically significant. On the third GR both
performed significantly better than other students in the course
($t_{63}=2.16, p<.05$). Up to this point the scores of the two groups did not
differ significantly from each other. On the final exam, however, Group 1's
continued improvement resulted in scores that were again significantly better
than the course average ($t_{40}=1.87, p<.05$). Group 2's performance declined so
that it was actually below the course average ($t_{20}=-0.68, n.s.$) and
significantly worse than that of Group 1 ($t_{63}=2.50, p<.02$).

Figure 1. Examination Results
Compared to Population Average

382
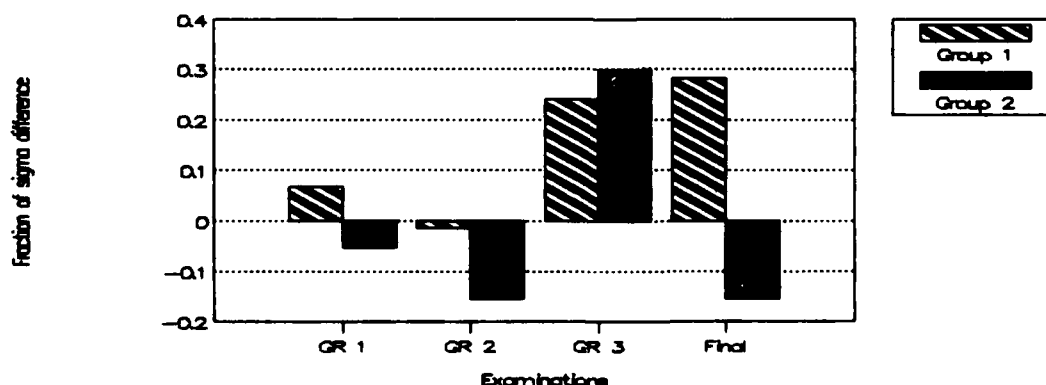
The results of the end of course survey are shown in Figure 2. Students in Group 1 viewed the course as more important ($t_{41}=2.51,p<.02$) while students in Group 2 viewed it as less important ($t_{19}=-1.90,p<.10$) than other students in the course. The difference in attitudes between the two groups toward the importance of the course was significant ($t_{86}=6.31,p<.001$). The two groups also differed significantly in their reported enjoyment of the course ($t_{86}=2.88,p<.01$) and their view of the subject matter ($t_{86}=3.69,p<.001$). The two groups reported similar attitudes about their instructor, the extra help they required, and the challenge of the course which were not significantly different from the course average.

Figure 2. Survey Results
Compared to Population Average
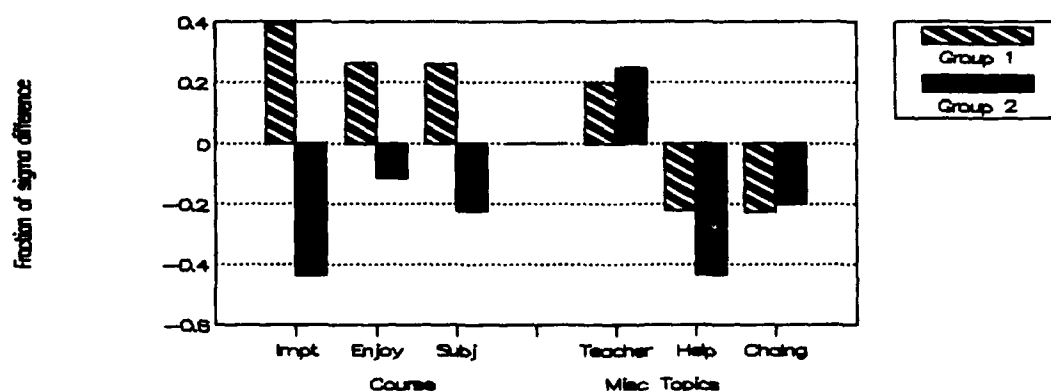
One survey question not shown on the figure asked the students if they thought doing computerized homework outside of class was a good idea. This was in reference to an effort by the Physics department to give points for correctly solving homework problems on the students own microcomputers in their rooms. They were allowed to get any help they needed from their peers. Group 1 responded more positively to this idea ($t_{41}=2.46,p<.02$) than other students in the course.

Student comments about the course were revealing. 51 of 64 students responded to the request for a critique by sending comments using electronic mail. Forty made a comments concerning cooperation or interdependent grading. Out of 49 comments, 42 were positive and 7 were negative.

The comments regarding cooperation were overwhelmingly positive. Students liked learning from their peers: "The whole class seemed to cooperate well together and worked to help people who had lower grades", "everyone was able to pool from their classmates knowledge ... an important factor in our success", "teamwork ... helped me learn a great deal more", "more of an Air Force-type of environment", "I liked the idea that we could work together without fear of being penalized for them getting good grades", and "personally I learned a lot by trying to teach others." They also liked being given more responsibility and appreciated the opportunity to control their own learning

and show leadership. However, a couple felt discomfort in being given more responsibility since "most of us couldn't come up with an original idea dealing with physics in a million years ... we need more direction", but this attitude was rare.

Regarding interdependent grading, there was some uneasiness about low performers relying on others' efforts and losing points because of others poor performance but, again, that was the exception. Most comments were positive and talked about the grading scheme as motivating them to work together.

One interesting thing that came out in the responses that was missed during the course of the semester was that a clique of athletes had developed within Group 2 that others in the group resented. This seemed to destroy group cohesiveness. Comments revealed that "intercollegiate athletes were treated much different [sic]", that there were "too many serious cliques that separated the people of the class", and that "it also built a lot of hate and resentment toward certain individuals".

Discussion

Results for Group 1 are generally consistent with our previous unpublished experience: significant positive shifts in both performance and attitudes. Results for Group 2, which still performed as well as the rest of the course, were disappointing but enlightening. They suggest that greater student interaction, responsibility, and control is a two edged sword. In a cooperative class, the development of exclusive cliques can negatively effect both attitudes and performance. Perhaps if I had played a more active role in teaching the "collaborative skills students need" (Johnson and Johnson, 1986), the group's problems could have been recognized and resolved. Collaborative learning techniques offer tremendous pedagogical opportunities. However, pitfalls, such as the formation of cliques, can exert a negative influence and reverse some of collaborative learning's demonstrated advantages.

References

Bruffee, K. A., "The Art of Collaborative Learning", Change, March/April 1987, 42-47.

Chickering, A. W. and Gamson, Z. F., "Seven Principles for Good Practice in Undergraduate Education", American Association for Higher Education Bulletin, March 1987, 3-7.

Johnson, D. W. and Johnson, R. T., "Computer Assisted Cooperative Learning", Educational Technology, January 1986, 14.

Kohn, A., "How to Succeed Without Even Vying", Psychology Today, September 1986, 22-28.

Wood, F. R. and Flannery, J. P. (1986). "A Critical Evaluation of the Competition System at the United States Air Force Academy in 1980", USAFA-TR-86-4, United States Air Force Academy, Colorado Springs, CO 80840

Cooperative Learning: The Army's New Approach for Training

Equipment Records and Parts Specialists[1]

Judith E. Brooks, Ph.D.
US Army Research Institute for the
Behavioral and Social Sciences

Abstract

Cooperative learning was implemented during an advanced individual training (AIT) course for Army Equipment Records and Parts Specialists. Course performance and administrative data were compared for cooperative and individual learning students to assess the usefulness and feasibility of group learning as a training technique. The results showed that cooperative learning incurred very little cost, was feasible to implement, and yielded some cost benefits, including a reduced academic recycle rate. Except for the first of six achievement tests, group learners performed as well as individual learners.

Cooperative learning methods are being investigated as one approach to increasing training effectiveness in the Army. The cooperative learning research reported in this paper was conducted at the Fort Lee Quartermaster School (QMS) as part of a Training Technology Field Activity (TTFA) program of research. The goal was to evaluate the feasibility and usefulness of cooperative learning as a training technique in the Advanced Individual Training (AIT) course for Equipment Records and Parts Specialists (MOS 76C).

Under cooperative learning, students spend part of their class time in small groups helping one another learn. This contrasts with conventional training approaches that require students to learn individually in the classroom with help from instructors but not from peers. According to Slavin (1983), two elements that are key to the success of cooperative learning are individual accountability and a clear incentive for the group to perform well as a whole. A preliminary investigation of cooperative learning in the Equipment Records and Parts Specialist Course by Hagman and Hayes (1986) indeed showed that individual soldier achievement was enhanced when students completed practical exercises (PEs) in four-member groups under a group incentive condition.

This paper describes a follow-on effort which was an extended nine-week tryout of Hagman and Hayes' cooperative learning method (see Brooks, Cormier, Dressel, Glaser, Knerr, & Thoreson, 1987 for a detailed account of this research). In this tryout, four-member cooperative learning groups under group incentive were compared with students who studied individually throughout the course. Both performance data and administrative data were gathered to more precisely determine the feasibility and usefulness of employing cooperative learning as a training method.

Method

Subjects and Design

The experiment was a between-subjects design. Three classes of 76C AIT students were randomly assigned to a cooperative learning condition, and three classes were assigned to an individual learning condition. Each class had a different instructor team, and the assignment of instructors to classes was random.

Procedure

In the cooperative learning classrooms, instructors assigned students to four-member study groups that had an even mix of high, average, and low performers. Starting with the second course annex, cooperative learning students completed daily PE assignments in their groups, formulating answers as a team. Group members were encouraged to help one another learn and to resolve questions about the material within the group, but could also request and receive help from an instructor if necessary. Each student had his or her own booklet for recording PE solutions, but since all of the answers within a group were the same, instructors needed to check only one PE per group for scoring and feedback purposes. Students in the individual learning condition completed all PE work individually, with occasional help from an instructor if required.

In both conditions testing was conducted individually, and each student was responsible for passing each end-of-annex exam before proceeding to the next annex. However, under cooperative learning, students were also accountable for the performance of their group members. If any member of the group failed an exam, all group members were required to attend a remedial study hall to help the failing member(s) study for a retest. Thus, there was an incentive for ensuring that each group member could pass the exams.

The cooperative learning procedures were implemented routinely for six annexes. In the seventh annex, which included a lengthy, comprehensive PE, students had an option to either continue working in groups or to complete the PE individually. At the end of the course, students and instructors also had an opportunity to express their opinions about cooperative learning by responding to a questionnaire.

Results

Achievement Test Scores

A 2 (learning condition) x 6 (annex) mixed analysis of variance (ANOVA) performed on the annex test scores yielded no main effect of learning condition but did yield a significant interaction between learning condition and annex, F (5,1170) = 6.33, $p < .05$. A Tukey post hoc test showed that individual learning students performed significantly better on the first annex test. There was no significant difference between learning conditions on final comprehensive test scores. Thus, the achievement test data suggested that group learners performed as well as individual learners, with the exception of the first annex exam.

Academic Recycle Rate

Academic recycling occurred in both learning conditions whenever a student twice failed a given annex test. Under cooperative learning, 4.4% of the students were recycled during the course, compared to a 10.9% recyle rate for individual learning students.

PE Performance

A 2 (learning condition) x 5 (annex) mixed ANOVA on the number of PE errors yielded a significant main effect of learning condition, F (1,228) = 223.21, $p < .05$. Cooperative learning students made fewer errors compared to individual learning students at all annexes. A similar analysis of PE completion times showed, however, that groups overall took significantly more time than individuals to complete PE assignments, F (1 229) = 35.69, $p < .05$. The learning condition x annex interaction was also significant, F (4, 916) = 44.78, $p < .05$, and Tukey post hoc tests showed that cooperative learning students took more time in early annexes, but actually took significantly less time during the last annex. Even when groups took longer to finish PEs, they did not exceed scheduled time limits.

Study Hall Data

Study hall records were examined to determine the impact of the cooperative learning approach with its group incentive on the study hall system. As anticipated, group learning study halls had a higher average attendance. An unexpected result, however, was that the instructors in the individual learning condition held more than twice as many study hall sessions; consequently, the average number of study hall hours per student was slightly higher under individual learning.

Attitude and Opinion Data

When given a choice near the end of the course to continue working in groups or to work individually, 92% of the students chose group PE work. Responses to informal questionnaires also indicated that most students and instructors liked cooperative learning.

Discussion

There were two notable benefits of cooperative learning. One was a savings of PE-scoring time for instructors. The other was a reduction in the academic recycle rate, which suggested that cooperative learning may increase the performance level of students who tend to perform most poorly in the course. Among students who were not recycled, group learners performed as well as individual learners on achievement tests overall. Contrary to expectation, though, group learners actually scored lower on the very first annex test, a result that may reflect students' need for time to adjust to the novel group learning procedures.

With respect to the PE work itself, the results showed that cooperative learning improved the overall quality of PE solutions. The data also showed that the lower PE error rate was accompanied first by an increase, then by a decrease in the time needed to complete PEs in a group setting. It appears that the group strategy that resulted in higher quality PE products also became more efficient over time as students gained experience working in groups.

An important finding for course administrators was that group PE work and the group incentive did not dramatically increase study hall attendance. Although cooperative learning did result in more students attending per session, the increase was a modest one that did not pose scheduling conflicts. In fact, the total number of study hall hours per student was higher in the individual learning condition, since the instructors in this condition unexpectedly held more sessions. The choice by control condition instructors to hold more sessions suggests that the tryout may have engendered competition among instructor teams. The study hall data indeed raise the question of whether any improvements in test performance due to cooperative learning were masked by the relatively greater training time associated with individual learning.

It was concluded that the cooperative learning approach used here is a low-cost and feasible training approach that offers potential benefits for the 76C course. Based on these findings, the Quartermaster School is employing cooperative learning in several 76C classes, while continuing to monitor its effects on course achievement and study hall requirements.

Finally, future research is recommended to identify positive group reinforcers that could be used as incentives for good test performance in lieu of the study hall incentive. The original decision to use study hall as a negative consequence of exam failure was driven only by its successful use in the earlier investigation (Hagman & Hayes, 1986) and by the constraints of the research setting. One alternative that is recommended for future evaluation is the use of between-group competition. Another avenue for future study is the use of specific group learning strategies that can be effectively used for training 76C tasks in manual and automated task environments.

References

Brooks, J. E., Cormier, S. M., Dressel, J. D., Glaser, M., Knerr, B. W., & Thoreson, R. (1987). Cooperative learning: A new approach for training equipment records and parts specialists (MOS 76C). (Technical Report No. 760). Alexandria, VA: US Army Research Institute for the Behavioral and Social Sciences.

Hagman, J. D., & Hayes, J. F. (1986). Cooperative learning: Effects of task, reward, and group size on individual achievement. (Technical Report No. 704). Alexandria, VA: US Army Research Institute for the Behavioral and Social Sciences.

Slavin, R. E. (1983). When does cooperative learning increase student achievement? Psychological Bulletin, 94, 429-445.

A Typical Day in the Life of Midshipmen and Their Attitudes about Training at the United States Naval Academy

Leanne E. Atwater Ph.D.
United States Naval Academy

Abstract

This study investigated how midshipmen at the U. S. Naval Academy use their time, and their attitudes about their education and training experiences. Time budget surveys, which asked midshipmen to record the amount of time spent in 60 activities on a specified day were obtained from 3785 midshipmen. Midshipmen described time spent in classes, studying, socializing, eating, sleeping, and many other activities. Attitude surveys were also received from 1870 mids. Results indicated that on the average, midshipmen spend about 3 1/2 hours a day on the weekdays studying, and about 2 1/2 hours relaxing, socializing, and recreating. Their attitudes about their experiences at USNA were generally positive. Remaining academically satisfactory and being in good shape physically were both very important to the mids. Those things that the mids felt were an advantage of USNA over a comparable civilian school they could have attended were getting a good education and learning self discipline. The lack of social development and the rules and regulations were seen as negative aspects of attending USNA. Overall, it was a useful survey for decision-makers at USNA and the instrument is available for others who may be interested.

The Naval Academy is continuously working to provide the best training and education for the select group of individuals involved in their program. Because of the unique features of USNA as an educational and training institution, all factors of a midshipman's life must be considered in attempts to achieve this goal. In the Spring of 1987, a study was done to assess how midshipmen spend their time on a typical day at USNA, and how they felt about their training experiences. This information has made a valuable contribution to efforts in self-assessment and has provided insight to decision-makers at the Academy. This technique is recommended as a useful tool for any institution concerned with time usage.

Method

Time-budget questionnaires were administered to all midshipmen in the brigade during the first week in April, 1987. Each midshipman was asked to provide a detailed account of his/her activities on one of the seven days of that week. In addition, a random sample of the brigade was asked to complete a

survey assessing attitudes about their training experiences. Completed time budgets were received from 3785 midshipmen and attitude surveys were received from 1870 randomly selected midshipmen. The surveys adequately represented each class.

Questionnaires

The time budgets asked mids to report in minutes the amount of time they had spent in 60 categories. The categories included such activities as athletic practice, listening to lectures in various classes, cleaning their room, thinking or relaxing, sleeping, studying, and many others. Mids were instructed to record their time usage at approximately four-hour intervals. Time budgets were completed by 2740 mids on weekdays and by 945 mids on either Saturday or Sunday.

The attitude surveys expanded on the time usage concept. Mids were asked how likely they would be to use liberty time to do various things (such as study for an exam, or engage in extracurricular activities). They were also asked how many hours they had available to study if they chose to, how much time they actually spent studying, and how important various aspects of Naval Academy life were to them. The attitude questions asked them to evaluate the importance of a number of things, including getting good grades, getting easy professors, making day to day decisions for themselves, getting away from the Academy, learning something interesting, being respected as "squared away" and other questions of this type. The attitude questions were responded to on four-point scales. The questionnaire also included two open-ended questions assessing what the mids felt were advantages and disadvantages of USNA compared to another college they could have attended.

Results

Time Budgets

While the results of the time budget study were very interesting and informative, they must be considered in light of a few methodological constraints. First, two of the days during the survey were part of an exam week which may have effected both the reports of time usage on those days as well as the weekend days following them. Second, the survey was conducted in April, just four weeks from the end of classes. This data is not representative of the ways many midshipmen (especially plebes and first class) spend their time at the beginning of the academic year. Third, it must be considered that this is self report data which is subject to errors and omissions, especially from otherwise busy individuals. Questionnaires were anonymous and it was specified that information obtained was for research purposes only, not to be used to tighten their schedules. For the most part, it appeared that the mids took the questionnaires seriously and tried to be accurate in their reports.

The distribution of time across a number of categories was interesting. One important category was studying. How much time do the midshipmen spend studying on an average weekday? When studying for all subjects are considered together, mids spend a little over 3.5 hours (average 214 minutes for 2367 cases) per weekday studying. The largest amount of time was spent studying engineering (average 128 minutes for 1051 cases) while the smallest amount of time was spent studying professional topics such as leadership and seamanship (average 87 minutes for 619 cases). (The number of cases are those persons who reported spending any amount of time in that category on the day questioned.) This estimate of study time coincides well with the estimates given in the attitude survey which will be discussed later in this paper.

Another activity of interest is sleeping. On the average, mids reported getting almost 7 hours of sleep on the weekdays (average 410 minutes, primarily at night). That average went up to almost 9 hours (530 minutes) per day on the weekends. These averages did not include sleeping in class which a small proportion of the mids admitted to having done. Roughly 15% of the mids reported sleeping for an average of 10-20 minutes in at least one class during the day.

Recreation, including TV, thinking, relaxing and socializing took up an average of 2.5 hours of the midshipmen's time on an average weekday. Cleaning their room averaged 23 minutes a day, with greater time being spent on inspection days. Time spent eating (all meals and snacks combined) averaged 1 hour per day. Two-hundred and seventy-three mids averaged 1.25 hours spent on extra curricular activities per day.

Time spent on the weekends was also of interest. Because each class rates different weekends, the weekend for the purposes of the time budgets was defined as Saturday and Sunday. Two-thirds of the midshipmen reported studying on Saturday or Sunday. Since all weekends technically end Sunday evening, the mids, particularly those on weekend liberty, often did their "weekend" studying on Sunday night. Not surprisingly, more studying was done by the 3rd and 4th class than by the 1st and 2nd class mids. The attitude surveys reported later in this paper lend additional insight into the attitudes about studying on the weekends. Roughly 25% of the mids questioned on Sunday reported engaging in religious activities.

Group differences in time usage were also of interest. In general, there were no differences between academic majors in the way time was spent, other than the amount of time spent studying for particular classes, (e.g. engineering majors spent more time studying engineering than did political science majors). This is not surprising in that a large number of core courses are required of all majors. When those with satisfactory cumulative grade point averages (>2.0) were compared with those with low cumulative grade point averages

(<2.0), the differences of note were that those with low averages were more likely to use their weekend time to study and spent less time relaxing and socializing than their academically satisfactory counterparts. There were no appreciable differences in the ways males and females spent their time.

Attitude Surveys

Midshipmen were asked to rate 28 aspects of their life at the Academy in terms of importance. These factors concerned grades, professors, athletics, learning, social activities, and professional activities. It was encouraging to learn that midshipmen felt getting good enough grades to keep a 2.0 grade point average was very important (mean 1.2; 1=Very Important). Also very important were being in good shape physically and learning things that will help them perform better as naval officers. Getting good/interesting professors was more important than avoiding professors known for strict grading policies. Of least importance (2.7; 3=Slightly Important) were getting good enough grades to please their parents and being respected by professors because they were bright. In general, mids were not particularly interested in "getting the gouge" for exams and avoiding hard work. They seemed quite motivated professionally, athletically and academically.

Few attitude differences emerged between male and female midshipmen. Females did, however, rate being respected as a "squared away" midshipman as more important than did males. Females were also less interested in getting the "gouge" for exams and avoiding strict professors than were males.

Also addressed on the survey were attitudes about using their weekend liberty time. Eighty-five percent of the mids questioned said they would be likely to use their weekend liberty time to study for an exam on Monday. Only 39 percent indicated that they would be likely to use their weekend liberty time to do an ordinary assignment for Monday.

One question of particular interest in this study asked midshipmen to estimate how many hours they had available to study if they chose to. The 1st 2nd and 3rd class mids estimated about 32 hours a week, (not including their weekend liberty time) were available to them for studying. The 4th class estimated 27 hours were available. When asked how much of this time was productively spent studying, estimates ranged from 68 percent of that time, or 18 hours by the plebes to 50 percent, or 15 hours by the 1st class. These estimates closely approximate the reports from midshipmen from all classes on the time budgets which averaged 3.5 hours of studying per day during the week (or 17.5 hours).

Two open-ended questions were included in the survey to assess what midshipmen felt were advantages and disadvantages of attending the Naval Academy as opposed to a civilian college.

The things included in their lists of advantages of USNA were getting a good education and learning self discipline and responsibility. The disadvantages listed included the lack of social development and the rules and regulations. It was interesting that learning a sense of responsibility for one's actions was an advantage yet the lack of responsibility for some of the basics like cooking and laundry was a disadvantage.

Discussion

Life at the Naval Academy is highly structured. Optimal allocations of time for various types of activities is important. Also of concern are the varying perceptions of faculty and staff as to just how much time midshipmen have to engage in various activities. While much speculation goes on, most viewers see only one portion of a midshipman's daily activities. The time budgets and attitude surveys provided a snapshot o⊥ how midshipmen _actually_ spend the day, and anonymously questioned their feelings about their experiences at the Academy. This study dispelled to some extent the belief that mids have a "2.0 and go" attitude and that they are so time constrained and sleep deprived that they are often looking for the easy way out. This study also raised questions about productive study, and how (or perhaps whether) study time can be used more effectively.

It would be of interest in the future to be able to look at similar data from the other academies and from comparable civilian institutions in order to assess similarities and differences, and perhaps to shed light on what is "normal" "necessary" or "optimal". For example, how much leisure time do these young people require in order to function at their best? Can we, should we, or do we need to change some of the midshipmens' attitudes about using weekend liberty time? All in all, this was a very useful and enlightening exercise for USNA. It is recommended to others as a way to find out "what cadets or students are really doing". Instruments used in this study are available for use by anyone interested.

Computer Contributions to Classroom Instruction

Gail A. Fallon, M.S.
David Porter, Ph.D.
Bruce R. Jaeger, Ph.D.
USAF Academy

Abstract

One lower division and two upper division psych-
lology courses were used to compare student percep-
tions as to what various instructional methods con-
tributed to their learning experience. Each method
was found to contribute differently to personal
enjoyment, critical thinking, and subject knowledge,
while showing similar trends across all three
courses. Computers were consistently rated highest
in personal enjoyment, followed by lectures, with
texts lagging far behind. For critical thinking,
all three instructional methods received similar
ratings. Lectures consistently rated highest for
subject knowledge, followed by texts and computers.

Educational research has been inconsistent in its findings
on the advantages of one educational method over another--this is
particularly true in the comparison of computer based instruction
to conventional classroom instruction. The reason for the incon-
sistency seems to be the inability to hold constant the quality
and the content of instruction across the various methods. Even
if the experimenter were able to hold content constant, it is
often at the sacrifice of some other aspect inherent to the
methods. Indeed, holding instructional content constant in a
conventional classroom setting severely hampers what may be its
greatest advantage (i.e., content flexibility). This study does
not attempt to find which method is "the best", but rather to
examine student perceptions of how each method contributes to the
course.

One of the most consistent findings in the literature on
computer based instruction (CBI) is that students seem to prefer
taking a course by this particular method. In studies of mili-
tary technical training, students enrolled in CBI courses rated
them more favorably than students enrolled in the same courses
taught by conventional methods (Dallman et al., 1977 and Longo,
1972). The findings on the civilian side, while not quite so
clear still show that students favor CBI over other methods. A
meta-analysis by Banget-Drowns et al. (1985) of eleven studies of
secondary students' attitudes found an average effect of .09 for
CRI over conventional classroom instruction, while a meta -
analysis by Kulik et al. (1980) of five studies of college stu-
dents' attitudes found an average effect of .18 favoring CBI.

Jones et al. (1983), and Avner et al. (1980) compared student ratings on how helpful CBI was in learning the material to other instructional methods: textbook, lecture, discussion section, quizzes, and laboratories. In both cases, CBI was rated as being the most helpful with the textbook the least.

While the Jones et al. (1983) and Avner et al. (1980) studies did make direct comparisons among the various instructional methods that the meta-analyses did not, they only compared the methods on a global scale of "helpfulness". This study seeks to see how students perceive each instructional method contributes to their learning experience in terms of three separate outcomes: personal enjoyment, critical thinking ability and knowledge of the course material. Each of these dimensions is considered to be an important criterion for evaluating course viability.

Method

Subjects

The subjects were all United States Air Force Academy cadets. There were 11 cadets enrolled in the cognitive psychology course, 64 cadets enrolled in the engineering psychology course, and 612 cadets enrolled in the general psychology course. The cognitive psychology course is an elective course for majors and was made up of juniors and seniors. The engineering psychology course is required of all human factors majors and was made up of seniors. The general psychology course is required of all cadets and was taken by freshmen. Twelve percent of the cadet population is women and the groups studied also approximated that proportion.

Materials

A department-wide course critique consisted of 31 statements about instructor performance with a 5-point rating scale ranging from "strongly agree" to "strongly disagree", followed by an

Figure 1. Outcomes matrix used by USAFA cadets to rate the contribution of each instructional medium.

"outcomes matrix" requiring cadets to rate each instructional method (see Figure 1) according to how it contributed to their personal enjoyment, ability to think critically, and knowledge about the course subject.

Design and Procedures

The survey was administered by instructors on the last day of class in each course. Cadets were asked to key in their ratings on a computer answer sheet and were instructed to consider only computer exercises and labs in responding to questions 10B, 15B, and 20B. All ratings were anonymous.

No attempt was made to control for the quality or quantity of instructional methods used in each course. Computers were used extensively in general psychology in the form of a computerized study guide and research demonstrations. Cadets used the computers in their rooms during study hours in this case. Computers were used for in-class demonstrations and experiments for the engineering and cognitive psychology classes. In engineering psychology, four lessons were reserved for computer demonstrations of common research paradigms in the field (e.g., signal detection, speed-accuracy tradeoff and reaction time). Six lessons were used for computer exercises in cognitive psychology.

Results

The mean ratings of each instructional method on personal enjoyment (PE), critical thinking (CT), and subject knowledge (SK) are given in Tables 1-3 below. A score of 4.0 would represent unanimous agreement by all students that the respective method contributed "a great deal" to a particular outcome (3.0: some; 2.0: moderately; 1.0: very little; 0: none).

Method	PE	CT	SK
Instructor	3.5	3.2	3.6
Lectures	2.9	2.9	3.2
Text	2.4	2.7	3.2
Films	3.2	2.6	2.9
Tests	1.4	2.6	2.6
Projects	2.0	2.3	2.4
Computer	2.9	2.9	3.0
Discussion	3.2	3.1	3.1
Overall	3.1	3.0	3.2
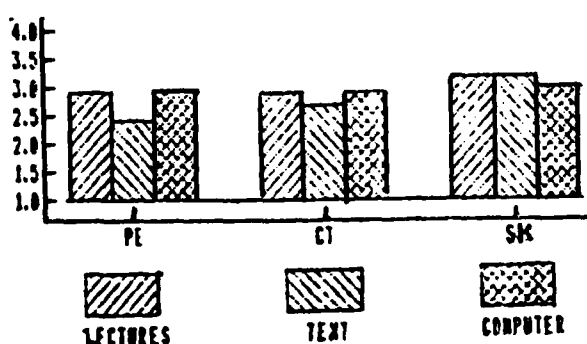
Table 1. Mean ratings of instructional methods in General Psychology. The computer rating reported is for PsychSim only.

Method	PE	CT	SK
Instructor	3.0	3.0	3.5
Lectures	2.2	2.5	3.1
Text	1.3	2.1	2.7
Films	*	*	*
Tests	1.4	2.3	2.5
Projects	2.1	2.5	2.6
Computer	2.5	2.4	2.5
Discussion	2.5	2.4	2.3
Overall	2.3	2.7	2.8
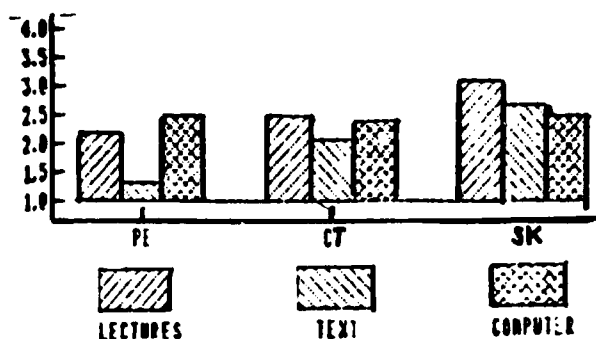
Table 2. Mean ratings of instructional methods in Engineering Psychology.

Method	PE	CT	SK
Instructor	3.6	3.9	3.9
Lectures	2.6	3.4	3.4
Text	1.3	3.1	3.3
Films	3.4	2.6	2.7
Tests	1.5	3.2	3.0
Projects	2.6	3.6	3.6
Computer	3.5	3.5	3.1
Discussion	3.4	2.8	3.1
Overall	2.9	3.7	3.6
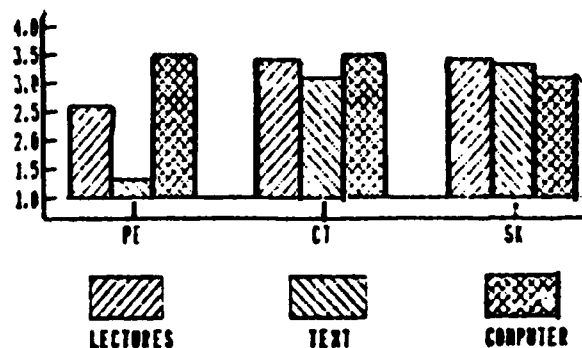
Table 3. Mean ratings of instructional methods in Cognitive Psychology.

Discussion

The three courses used for the study vary on many dimensions. In fact, their variety in enrollment, level and overall student ratings provide a general representation of the behavioral science curriculum at the Air Force Academy. This variance between courses, coupled with the different ways in which computer exercises were used within each of the courses, makes the similarities in rating patterns particularly interesting.

In each course "the instructor" was identified by students as contributing the most to their personal enjoyment, critical thinking ability and subject knowledge. These positive ratings reflect both the high caliber of the USAFA faculty and the special emphasis on classroom teaching by the department. After the instructor, computer exercises were rated as the most enjoyable aspect of two of these courses. In fact, students' ratings of computer exercises were very similar to their ratings of films and group discussions in all three courses.

Students' ratings of the contribution of computer exercises to their critical thinking ability in both general and engineering psychology was near the middle of the eight other instructional methods. In cognitive psychology, it was rated a close third behind a major research project. It appears that students' perceptions of the contribution computer exercises make to their

critical thinking ability depends upon the specific computer exercise and course context.

In all three courses, the computer exercises were rated slightly lower than other methods in their contribution to students' acquisition of course knowledge. This probably reflects the fact that in all three courses computer exercises were used to augment and supplement the primary academic material rather than to simply reiterate important points. In spite of this, students' overall opinion was that the computer exercises made "some" (2.8) contribution to the acquisition of knowledge.

Although student ratings vary from course to course on how much the computer contributes to critical thinking ability and course knowledge, computers are consistently rated relatively high on their contribution to personal enjoyment. This suggests that earlier studies which used a single, global criterion may have been influenced more by students' enjoyment of the computer exercises than by their substantive contributions to students' critical thinking ability or knowledge of the subject. Related research is being planned to incorporate objective dependent measures of both critical thinking and course knowledge as well as the subjective student ratings used in this study.

References

Avner, R.A., Moore, C., & Smith, S. (1980). Active external control: A basis for the superiroity of CBI. Journal of Computer-Based Instructions, 6(4), 115-118.

Bangert-Drowns, R.C., Kulik, J.A., & Kulik, C. (1985). Effectiveness of computer-based education in secondary schools. Journal of Computer-Based Instructions, 12(3), 59-68.

Dallman, B.E., DeLeo, P.J., Main, P.S. & Gillman, D.C. (1977). Evaluation of PLATO IV in vehicle maintenance training (Report No. AFHRL-TR-77-59). Chanute Air Force Base, IL: USAF School of Applied Aerospace Sciences, Training Research

Jones, L.M., Kane, D., Sherwood, B.A., & Avner, R.A. (1983). A final-exam comparison involving computer-based instruction. American Journal of Physics, 51(6), 533-538.

Kulik, J.A., Kulik, C., & Cohen, P.A. (1980). Effectiveness of computer-based college teaching: A meta-analysis of findings. Review of Educational Research, 50(4), 525-544.

Longo, A.A. (1972). A summative evaluation of computer assisted instruction in U.S. Army Basic Electronics Training (Technical Report No. 72-1). Fort Mnouth, NJ: US Army Signal Center & School.

The Incidence of Bulimia Among Female Cadets
at the United States Air Force Academy

Heidi Terhune
Carl Bryant, Ph.D.
USAF Academy

Abstract

This study was an attempt to replicate the earlier work of
Pyle, et al (1983) in the area of assessing the incidence of
bulimia in college women. The specific purpose of this study was
to determine the incidence of bulimia among female cadets who
attend the Air Force Academy and to compare the results with
findings from other college populations (Halmi et al, 1981, and
Pyle et al, 1983). It was the researcher's hypothesis that,
based on the competitive Academy environment and perceived
pressure for perfection, the incidence of bulimia would be higher
for female cadets than for other college females. Significant
differences were found between female cadets and women attending
civilian universities. In addition, significant differences were
found between classes of female cadets.

Bulimia is a relatively new disorder which has only received attention
since the late 1970s. It has been described as a devastating,
all-encompassing, sometimes life-threatening disorder with similarities to
chemical dependency (Cauwels, 1983). For a period of time, it was considered
to be a complication of anorexia nervosa (Abraham & Beumont, 1982; Russell,
1979) even though the American Psychiatric Association gave it a separate
classification as early as 1980. The diagnostic criteria for this disorder
are listed in Table 1.

The characteristics of this disorder have been described in numerous
articles and books (Abraham & Beumont, 1982; Halmi et al., 1981; Leon, 1983,
Pope & Hudson, 1984; Pyle et al., 1981; Russell, 1979). The work of these
authors has been summarized as follows: Bulimia consists of an uncontrollable
urge to ingest inordinately large amounts of high caloric, easily ingested
food. This portion of the disorder is defined as binging. The period of
binge eating lasts from minutes to hours and is frequently conducted as a
secretive act. The episode of binging usually terminates when the bulimic is
interrupted by social disturbances, abdominal pain, or the process of purging.
Purging has been described by the authors listed above as consisting of one or
all of the following characteristics: self-induced vomiting; use of
diuretics; use of large quantities of laxatives; or, less frequently,
excessive exercise. This study attempted to measure the degree these
characteristics were demonstrated by female cadets and possible differences by
academic class.

400

Method

Subjects

The subjects in this study consisted of 99 female cadets who were attending the United States Air Force Academy during the Fall Semester of 1987. This sample of 99 consisted of 27 freshmen, 26 sophomores, 21 juniors, and 25 seniors. Subjects were randomly selected from the entire female cadet population. In all, 148 surveys were sent to female cadets, and a total of 99 were returned. This accounted for a return rate of 66.9 percent.

Instrumentation

A modified version of the questionnaire developed by Pyle et al. (1983) was used to survey the subjects in this study. It contained 41 items which were constructed to gather information in six areas. These areas included: basic demographic data, incidence of bulimia relative to the DSM-III diagnostic criteria, frequency of using weight control methods, weight and food attitudes, body image, and history of treatment for eating disorders.

Results

The results of this study are reported in three sections. The first reports the incidence of bulimia; the second, differences in results by class. The third section reports the results of additional analysis.

Incidence of Bulimia

The responses to the items designed to measure the incidence of bulimia resulted in 12 cadets being labeled bulimic. These 12 cadets consisted of two seniors, 2 juniors, 2 sophomores, and 6 freshmen. These 12 "bulimic" cadets made up 12.1% of all the subjects who returned the survey.

Differences by Class

Analysis of variance procedures (ANOVA) were used to test for possible differences between academic classes. If significant differences were found, post hoc analyses were conducted using the Sheffe method as described by Thorndike (1982). Four variables were found to significantly differ by class. The first variable was frequency of binge eating. On this measure, juniors were found to binge sig. .ficantly less often than the other three classes surveyed. The second variable in which significant differences by class were found was the variable measuring fear of becoming fat. In this area, seniors were significantly less fearful of becoming fat than any of the other classes surveyed. On the variable measuring if a subject felt they constantly overate, freshmen were found to be significantly higher on this scale than any of the other classes. The final area in which significant differences by

class were found was on the variable which measures frequency of eating when under stress or feeling bad. Freshmen were found to be significantly higher on this scale than juniors or seniors. No significant difference between freshmen and sophomores was found.

Additional Analysis

While 12.1% of the subjects surveyed met the inclusion criteria for bulimia, the population overall showed some interesting results on the weight-related issues. Overall, the female cadets surveyed were at 84% of their maximum allowable weight (MAW) according to Air Force standards. Thirty-three percent of the subjects who were identified as "bulimic" were within 10% of MAW while only 14.9% of the nonbulimic subjects were within 10% of their MAW. Bulimics wanted to lose, on the average, 8.5 pounds or 6.4% of their body weight. Nonbulimics wanted to lose 7.2 pounds or 5.6% of their body weight. The 99 female cadets surveyed reported overwhelmingly (71.7%) that they had a fear of becoming fat and that 53.5% of them felt fat even though others disagreed with them. 68.7% of the subjects in this study said they ate when under stress or feeling bad and 60.6% admitted to binge eating.

Discussion

The first hypothesis, that the incidence of bulimia would be higher in the female cadet population than in the population reported by Pyle et al. (1983), was supported. In this study, an incidence of bulimia of 12.1% was found, while in the Pyle study, an incidence of bulimia of only 7.8% was reported. since all of the subjects in the Pyle et al. (1983) study were freshmen, perhaps it would be of interest to compare the results of that earlier study with only the freshmen in the current study. When this analysis was conducted, and even more marked difference between the studies was found with 22.2% of the freshmen in the current study falling into the bulimic category.

The second hypothesis, that there would be no significant difference by class, was rejected. Juniors were found to report binging less than the other classes, and seniors were less fearful of becoming fat. Freshmen were found to be far more likely to think they constantly overate and were more prone to eat when under stress or feeling bad. This data on freshmen appears to be critical in light of their higher incidence of bulimia.

The additional analyses which were conducted warrant further research. With over 70% of those surveyed reporting fear of becoming fat and over 50% of those surveyed stating they felt fat even when others disagreed, it would seem possible that female cadets are experiencing serious problems in the areas of weight control and body image. Additionally, with almost 70% of those surveyed reporting eating when under stress or feeling bad, it would appear that female cadets could benefit from classes directed at teaching more appropriate coping skills for stress. Additional research is needed to measure the possible impact of these actions.

TABLE 1

DIAGNOSTIC CRITERIA FOR BULIMIA

A. Recurrent episodes of binge eating (rapid consumption of a large amount of food in a short period of time, usually less than two hours).

B. At least 3 of the following:

    1. consumption of high-calorie, easily-ingested food during a binge

    2. inconspicuous eating during a binge

    3. termination of such eating episodes by abdominal pain, sleep, social interruption, or self-induced vomiting

    4. repeated attempts to lose weight by severely restrictive diets, self-induced vomiting, or use of cathartics and/or diuretics

    5. frequent weight luctuations greater than 10 pounds due to alternating binges and fasts

C. Awareness that the eating pattern is abnormal and fear of not being able to stop eating voluntarily.

D. Depressed mood and self-deprecating thoughts following eating binges.

E. Bulimic episodes are not due to anorexia nervosa or any known physical disorder.

References

Abraham, S. F. & Beumont, P. J. (1982). How patients describe bulimia or binge eating. _Psychological Medicine,_ _12_, 625-635.

American Psychiatric Association (1980). _Diagnostic and Statistical Manual of Mental Disorders_, 3rd ed., American Psychiatric Association, Washington, D.C.

Halmi, K. A., Falk, J. R., & Schwartz, E. (1981). Binge-eating and vomiting: a survey of a college population. _Psychological Medicine,_ _11_, 697-706.

Leon, G. R. (1983). _Treating eating disorders: Obesity, anorexia nervosa, and bulimia_. Lexington, MA: The Lewis Publishing Co.

Pope, H. G. Jr. & Hudson, J. I. (1984). _New hope for binge eaters: Advances in the understanding and treatment of bulimia_. New York: Harper & Row, Publishers.

Pyle, R. L., Mitchell, J. E., & Eckert, E.D. (1981). Bulimia: A report of 34 cases. _Journal of Clinical Psychology,_ _42, 60-64._

Russell, G. (1982). Bulimia nervosa: An ominous variant of anorexia nervosa. _Psychological Medicine,_ _9_, 429-448.

Thorndike, R. M. (1982). _Data collection and analysis: Basic statistics._ New York: Gardner Press, Inc.

The Impact on Academic Persistence Among Black Cadets at the Air Force
Academy as Measured by Noncognitive Variables

Kevin W. O'Callaghan, PhD
Carl Bryant, PhD
Raoul Buron, Jr., MA

Abstract

There has been a plethora of research conducted that has
addressed the attrition problems for black students attending
predominantly white colleges and universities. The present research
was conducted, in part, to provide descriptive and potential
correlational data as to the impact of noncognitive variables,
particularly among black students, on persistence in higher
education. The attrition rate for blacks at the Air Force Academy
is far below that of most other predominantly white institutions.
Thus, a second purpose of the study was to compare the entering
black and white freshmen at the Air Force Acadaemy to those at a
predominantly white state university to ascertain if there were
significant differences in the selection and admissions process.

The data was collected using the Sedlacek Noncognitive
Questionnaire (NCQ). Subjects in the study consisted of 122
randomly-selected undergraduates at the Academy (62 black and 60
white) in their freshmen and senior years. Data obtained from the
Sedlack NCQ was analyzed using ANOVAs and a t-test.

The results indicated significant differences existed between
entering black and white cadets at the Air Force Academy compared to
their civilian counterparts. Significant differences were also
found between groups at the Academy.

In the late 1970s and early 1980s, the issue of retention and attrition
of students in institutions of higher education became extremely important.
Pantages and Creedon (1978) stated that for all students who enter four-year
colleges each Fall, only 40 percent graduate within the traditional four
years. It has also been well documented that the attrition rate is much
higher among black students than it is among whites (Astin 1975, 1978, 1982;
Sedlacek and Pelham, 1976). This difference in attrition rates is magnified
when the institutions of higher education examined are those that are
predominately white (Goodrich, 1978).

Fleming (1984) stated that social scientists have amassed a considerable
body of literature indicating that all is not well with black students in
predominantly white colleges. She states, "The recurrent problems of alleged
hostility and racism, poor rapport with faculty, feeling left out of the
curriculum, inadequate social lives, exclusion from campus activities, the
conflict of a demanding black subculture, and the experience of academic
failure are said to be indicators of a crisis in social adjustment symptomatic

of a more serious identity crisis."

This crisis leads black students in predominantly white istitutions to feel increasingly dissatisfied with the formal and institutional aspects of the undergraduate experience. The end result of this feeling of lack of institutional support and isolation often leads to academic failure or voluntary attrition (Fleming, 1984). Research by Astin (1977) reveals that dropping out of a college or university is a problem for individual students because it denies them access to many higher-paid and higher-skilled jobs. This is especially true for minorities who, upon leaving school, display less intellectual self-esteem and competence and pooer interpersonal relationships (Astin, 1977; Chickering, 1981; Fleming, 1984; Sedlacek, 1982; O'Callaghan, 1986).

Little research has been conducted as to how black students have fared at a predominantly white four-year military academy. The attrition rate for black students currently attending the United States Air Force Academy (1987-1990) has averaged less than 20%. This is significantly below the average attrition rates for blacks at other predominantly white institutions. Thus, the purpose of this study was to explore why the attrition rate for black students at the Air Force Academy appears significantly lower than that of other predominantly white institutions by describing the response patterns of black and white college freshmen and seniors at the Air Force Academy and comparing them to those in previous research studies conducted by Sedlacek (1976) and Fleming (1984) using noncognitive variables.

Method

A cross-sectional, casual-comparative approach was employed in the research. Seven noncognitive variables were measured by the Sedlacek and Brooks Noncognitive Questionnaire. These variables were (1) positive self-concept; (2) understands and deals with racism; (3) realistic self-appraisal; (4) preference to long-range goals to short-term goals; (5) availability of a strong support person; (6) successful leadership experience; and (7) demonstrated community experience. Between August 1, 1986, and October 1, 1986, the NCQ was administered to a randomly-selected sample population.

The subjects were 122 cadets attending the United States Air Force Academy in Colorado Springs, Colorado. Their ages ranged from 17 to 25 years of age. Those sampled were black and white students from the freshmen and senior classes. The study sample consisted of 32 black and 30 white students from the freshman class and 30 black and 30 white students from the senior class. The participants in the study were instructed to complete the questionnaire that took under 30 minutes to complete. Subjects were told that participation in the study was voluntary and that there was not a time limit for completing the questionnaiure.

Data gathered relative to the NCQ was statistically analyzed using an ANOVA and t-test. The Statistical Package for Social Sciences (SPSS) was used to conduct the statistical analysis.

Results

The results indicated that the entering black and white freshmen at the

Air Force Academy were significantly different than their counterparts at the state university. The white students at the Academy scored significantly higher ($p < .01$) on all seven of the noncognitive variables than did their civilian counterparts while, for black cadets at the Academy, significance was found ($p < .01$) on six of the seven noncognitive variables. These results strongly indicate a significant difference in the selection process of the two institutions.

Significant differences were also found between groups at the Academy. Positive self-concept ($p < .017$) and the ability to identify and deal with racism ($p < .001$) were significant by race. In both cases, blacks scored higher than whites. Black freshmen were significantly higher ($p < .001$) than white freshmen on self-concept while white freshmen were higher ($p < .026$) on demonstrated community service. No significant differences were found between black freshmen and seniors on any of the seven variables. A significant difference ($p < .007$) was detected between black and white seniors on the racism variable.

Discussion

The results of this study detected a number of interesting findings. Some of these came as no surprise while others contradicted much of the current literature on the persistence of black students in higher education. The following conclusions were made:

(1) The selection process at the Air Force Academy appears to be significantly different for black and white students as compared to the state university looked at in this study. When viewing the low attrition rate for the Academy, it was evident that the quality of their students is unique. A closer look at the Academy's admissions process shows a high degree of importance is placed on many of the noncognitive variables analyzed in the study in addition to the traditional SAT/ACT scores and high school grades. Thus, the high retention rate for cadets at the Academy is of no surprise to the researcher.

(2) The lack of overall differences between and among groups studied at the Academy indicates that the cadets are fairly homogeneous and similar in their desires, goals, and academic preparedness. These results also suggest that the strong noncognitive support systems that Sedlacek (1978) and Fleming (1984) state as critical in retaining blacks in college are well established at the Academy. The important role of noncognitive variables in black student success at the Air Force Academy may be evidenced by the strong academic advisement programs, close faculty-student relationships, various and numerous leadership opportunities, and the knowledge that these students have definite careers ahead of them.

(3) The last general conclusion resulting from this study is that, while the Academy appears to be doing things well, there is still room for improvement. This is clearly evidenced by the noncognitive variable of dealing with racism. The findings indicated that black seniors saw this as a more significant factor than did the white seniors. It appears likely that the institution needs to closely monitor this potential area of concern. The lack of sensitivity on the part of white seniors in this study, if not dealt with seriously, could lead to a less productive college environment.

References

Astin, A. W. (1975). Preventing students from dropping out. San Francisco, CA: Jossey-Bass.

Astin, A. W. (1977). Four critical years: Effects of college on beliefs, attitudes, and knowledge. San Francisco, CA: Jossey-Bass.

Astin, A. W. (1978). Four critical years. San Francisco, CA: Jossey-Bass.

Astin, A. W. (1982). Minorities in American higher education. San Francisco Washington, London: Jossey-Bass.

Chickering, A. W. and Associates (1981). The modern American college. San Francisco, Washington, London: Jossey-Bass.

Fleming, J. (1984). Blacks in college. San Francisco, CA: Jossey-Bass.

Goodrich, A. (1978, March). A data-driven minority student retention model for faculty and administrators in predominantly white institutions. Paper presented at the annual convention of the American College Personnel Association, Detroit, MI.

O'Callaghan, K. W. (1986). The effects of noncognitive variables on academic persistence among blacks at a predominantly white institution. Paper presented at the 2nd National Conference on Black Student Retention in Higher Education, Atlanta, GA.

Pentages, T. J. and Creedon, C. F. (1978). Studies of college attrition: 1950-1975. Review of Educational Research, 48, 49-101.

Sedlacek, W. E. and Brooks, G. C., Jr. (1976). Racism in American education: a model for change. Chicago: Nelson-Hall.

Sedlacek, W. E. (1982). The validity and reliability of a noncognitive measure of minority student retention. Research Report #3-82, University of Maryland, College Park.

Sedlacek, W. E. and Pelham, J. C. (1976). Minority admissions to large universities: A national survey. Journal of Non-White Concerns in Personnel and Guidance, 4(2), 53-63.

The Relationships of Physical Attractiveness and Perceived Masculinity to the Performance Evaluation and Occupational Success of Naval Officers[1]

Virginia M. Berry, Ph.D.
Herbert George Baker, Ph.D.
Navy Personnel Research and Development Center

Abstract

This field study examined the relationship between a naval officer's physical appearance and his academic performance at the U.S. Naval Academy and subsequent occupational success. Results indicated that at the extreme levels of attractiveness, appearance is related to military aptitude ratings assigned in the academy and the rate at which an officer is promoted. For the majority of naval officers, appearance is not related to performance evaluation or occupational success.

The literature is replete with studies of the relationship between physical attractiveness and employment decisions. However, the vast majority of these studies have utilized undergraduate college students as subjects, bogus resumes, or fictional job applicants. There is very little field research in the area of appearance and performance evaluation and occupational success. The purpose of the current study was to determine whether physical attractiveness or a masculine appearance are related to academic performance in a military service academy and subsequent career success as a naval officer.

Method

Subjects

Subjects were 30 male naval officers at or above the rank of Lieutenant Commander and 30 male Masters of Business Administration students from two universities in San Diego, California. Participation in the research was voluntary and anonymous.

Materials

Materials consisted of 50 randomly selected black-and-white photographs (3 1/2" x 5") of 1973 U.S. Naval Academy graduates, and a set of 1 to 9 point rating scales with "not at all" and "very much" as endpoints. The rating scales were comprised of six questions (2 test items, 4 distractors): (1) How much does this person look like a leader? (2) How outgoing or extroverted does this person look? (3) How physically attractive is this person? (4) How intelligent looking is this person? (5) How masculine looking is this person? and (6) How much do you think you would like working with this person?

---

[1] The opinions presented in this paper are those of the authors and do not necessarily reflect official U.S. Navy policy.

Procedures

Subjects filled out a demographic information form and received brief training on how to avoid common rating errors. Subjects then assigned ratings the targets photographs in response to the six questions listed above.

Results

For each sample, the physial attractiveness and perceived masculinity judgments were correlated with four success measures: (1) cumulative grade point average (GPA), (2) cumulative military aptitude ratings (CMAR), (3) rank attained as of 1987, and (4) recommendations for early promotion (EPROM).

No statistically significant correlations were found between physical attractiveness or masculine appearance and any of the success measures (see Table 1). The lack of relationship occurred regardless of the subject sample (naval officer or MBA student) and duty status of the target (currently on active duty in the Navy vs. nonactive duty). Results obtained by combining the subject samples (N=60) were identical to those uncovered in the primary analyses.

Table 1

Intercorrelations of Appearance and Success Measures

Success Variable	Physical Attractiveness Naval Officers	MBAs	Masculine Appearance Naval Officers	MBAs
Grade Point Average				
Full Sample	-.03	-.05	-.17	-.14
Currently Active	-.22	-.20	-.34	-.33
Nonactive	.14	.12	-.05	.01
Cumulative Military Aptitude Rating				
Full Sample	.11	.10	-.06	.00
Currently Active	-.12	-.19	-.34	-.29
Nonactive	.25	.28	.12	.17
Rank Attained				
Full Sample	.06	.14	-.04	.01
Currently Active	----	----	----	----
Nonactive	.00	.04	-.08	-.02
Recommendations for Early Promotion				
Full Sample	-.01	.06	-.13	-.06
Currently Active	-.23	-.24	-.37	-.25
Nonactive	.06	.10	-.01	.03

Note. For the full sample, N=50; for currently active, n=23; and for nonactive, n=27. Correlations were not computed for currently active subset on the criterion rank attained due to lack of variability (all were Lieutenant Commanders).

In contrast to previous studies, the present study did not use target photographs that were prerated as extremely attractive or unattractive. In other words, the present study used the whole spectrum of attractiveness, not just the extremes. To determine whether this approach influenced the results, a more conventional extreme levels approach was simulated. The physical attractiveness and masculine appearance ratings of the 5 most attractive and the 5 least attractive targets were correlated with the success measures. Using the extreme levels of attractiveness in this manner produced results that were slightly different from those uncovered by the primary analyses (see Table 2). Physical attractiveness was related to recommendations for early promotion for the naval officer sample. Masculine appearance was related to recommendations for early promotion for the naval officers, MBA students, and both samples combined. Masculine appearance was also related to military aptitude in the academy for the naval officer sample and both samples combined, but not the MBA sample.

Table 2

Intercorrelations of Appearance and Success Measures for the
5 Most Attractive and 5 Least Attractive Targets

Success Variable	Physical Attractiveness	Masculine Appearance
Grade Point Average		
Naval Officers (N=30)	.21	.20
MBA Students (N=30)	.09	.19
Combined Samples (N=60)	.15	.20
Cumulative Military Aptitude Rating		
Naval Officers (N=30)	.55	.70*
MBA Students (N=30)	.46	.58
Combined Samples (N=60)	.51	.66*
Rank Attained		
Naval Officers (N=30)	.52	.48
MBA Students (N=30)	.58	.55
Combined Samples (N=60)	.56	.52
Recommendations for Early Promotion		
Naval Officers (N=30)	.64*	.69*
MBA Students (N=30)	.56	.73*
Combined Samples (N=60)	.60	.72*

*$p < .05$

Discussion

The results of this field study indicate that when the whole spectrum of appearance is considered, there is no relationship between appearance and performance evaluation or occupational success of naval officers. In other words, physical attractiveness and perceived masculinity play a small part, if any, in the promotion and career growth of naval officers. The data suggest, however, that at the extreme levels of appearance, such a relationship may indeed exist.

The present findings may be tied into research showing that the attractiveness bias operates only when women are being evaluated (Kaplan, 1978; Heilman & Stopeck, 1985). Unlike the majority of previous studies, in the present study men rated men, not women. As Kaplan (1978:202) stated, "the attractiveness halo effect is specific to situations in which men judge the low quality work of women." Results of the present study support the finding that the effects of a man's physical appearance are negligible when he is being evaluated by another man. The present study supports other research failing to confirm the relationship between appearance and performance evaluation or occupational success (e.g., Carlson, 1967; Sparacino, 1980).

A core explanation for the present findings is that it was one of the first field studies dealing with appearance and academic performance or occupational success. "Virtually all of the investigations completed to date have consisted of laboratory studies utilizing student raters. This approach has impeded the external validation of findings to more common evaluative contexts. That is, the raters typically employed have had little experience in rendering any type of formal interpersonal evaluations" (Morrow & McElroy, 1984:172). Thus, the present research may reflect the reality of employment better than previous laboratory research. Whereas previous studies used undergraduate college students and/or fictitious job applicants, the present study used the actual academic and professional data of the targets along with subjects who were similar to those who evaluate the targets in daily life. Under such "real life" conditions, the relationship between one's appearance and success (both academic and occupational) appears to be negligible.

Previous laboratory research utilizing extreme levels of appearance should be reevaluated in light of the present findings. Field research in the area of appearance and performance evaluation and occupational success is still needed. Aristotle once stated, "beauty is better than any letter of introduction" (Berscheid & Walster, 1974). However, as Cash (1981:18) points out, "in most respects Aristotle was right; but then again, in other respects, perhaps he should have collected more data."

References

Berscheid, E., & Walster, E. (1974). Physical attractiveness. In L. Berkowitz (Ed.), Advances in experimental social psychology. New York: Academic Press.

Carlson, R. E. (1967). Selection interview decisions: The relative influence of appearance and factual written information on an interviewer's final rating. Journal of Applied Psychology 51(6), 461-468.

Cash, T.F. (1981, August). The interface of sexism and beautyism. Paper presented at the annual convention of the American Psychological Association, Los Angeles.

Heilman, M.E., & Stopeck, M.H. (1985). Being attractive, advantage or disadvantage? Performance-based evaluations and recommended personnel actions as a function of appearance, sex, and job type. Organizational Behavior and Human Decision Processes 35, 202-215.

Kaplan, R.M. (1978). Is beauty talent? Sex interaction in the attractiveness halo effect. Sex Roles 4(2), 195-204.

Morrow, P.C., & McElroy, J.C. (1984). The impact of physical attractiveness in evaluative contexts. Basic and Applied Social Psychology 5(3), 171-182.

Sparacino, J. (1980). Physical attractiveness and occupational prestige among male college graduates. Psychological Reports 47, 1275-1280.

The Cognitive Development of Air Force Officer Candidates

Carl Bryant, Ph.D.
Kevin W. O'Callaghan, Ph.D.
Raoul J. Buron, Jr., M.S.
Gary L. Lorenzen, M.S.
USAF Academy

Abstract

This study investigated the cognitive developmental level of officer candidates using William Perry's Model of college student development. Officer candidates from five geographic regions and two service academies were surveyed using Moore's (1986) Learning Environment Preference Scale. Data were analyzed using analysis of variance procedures. Significant differences were found which related to commissioning source, academic class, and grade point average. Implications for teaching officer candidates are also addressed.

Recent research in the area of college student development has found that it is possible to measure cognitive development in college students using Perry's (1968) model. This model focuses on assessing a student's ability to handle increasing complexity and ambiguity in a learning environment. Extensive research by Mentkowski, Moeser, and Strait (1983) has found that, not only can cognitive development be measured, it can also be enhanced using specific methods of instruction. However, when the cognitive developmental level of students attending military academies was assessed by Bryant (1982), Phillips (1984), and Rice (1986) using the Measure of Intellectual Development (MID), lower than theoretically expected levels of development were found. In fact, Phillips (1984) cites the final report of the West Point Study Group (1977) as showing, ". . . shortcomings in dealing with issues for which there are no clear, right answers result from cadets having too few opportunities to study and solve problems characterized by ambiguity rather than certainty. Decision making in combat deals most often with uncertainty. The most successful wartime leaders have been trained to sort meager, often conflicting data to develop a workable solution when none is perfect and then to execute the plan well. . . Responses from the field appear to be saying that the curriculum does not adequately prepare cadets for such situations" (p. 69). The current study attempted to measure the cognitive developmental level of officer candidates who were enrolled in Air Force ROTC detachments nationwide and to compare their developmental level with those of cadets attending the Air Force Academy and a sister service academy in an attempt to measure whether officer candidates trained in civilian universities showed a significantly different developmental level than that of service academy cadets.

Method

The researchers determined that, in order to truly assess the cognitive developmental level of officer candidates nationwide, it was necessary to collect data from ROTC students who were attending college in five geographic regions. The regions were designated "North, South, East, West, and Midwest". The subjects consisted of 1045 officer candidates. The ROTC sample consisted of a total of 562 subjects. The Air Force Academy sample was designed to include 400 cadets. However, only 321 questionnaires were returned. The sister service academy student sample consisted of 162 subjects. All subjects were randomly selected from within their academic class. All the cadets sampled were asked to complete the fourth edition of the Learning Environment Preferences instrument (Moore, 1986). This instrument was designed to provide an objective, fast, and reliable method of assessing cognitive development along the Perry scheme. The data were analyzed using a pascal scoring program for the Learning Environment Preferences (LEP) which was developed by one of the authors. All data were then analyzed using analysis of variance procedures. These procedures were accomplished using the Statistical Package for the Social Sciences (SPSS).

In order to investigate possible differences related to academic ability, subjects in this study were categorized into five distinct groups based on GPA. The first group consisted of those students who had a GPA of 0.00-1.99. The second group had GPAs ranging from 2.00-2.49. The third group had GPAs of 2.50-2.99. The fourth group had GPAs from 3.00-3.49, and the fifth group had GPAs ranging from 3.50-4.00.

Results

The results of this study were divided into seven sections which included the following categories: Cognitive Complexity, Degree of Relativism, View of Learning, Role of the Instructor, Role of Students, Classroom Activities, and Grading Procedures.

Cognitive Complexity

Significant differences were found in the cognitive complexity of the subjects by source of commissioning and by grade point average (GPA). Differences were significant at beyond the $P < .001$ level. Significant interactions between source and class and between source and GPA were also found.

Degree of Relativism

Significant differences by GPA were also found in the percentage of relativistic thinking (Perry Position 5) that subjects selected ($p = .001$). A significant interaction between source and class were also found.

View of Learning

Significant differences were found by commissioning source ($P < .05$) and GPA ($P < .05$).

Role of the Instructor

No significant main effects or interactions were found for this variable.

Role of Students

No significant main effects or interactions were found for this variable.

Classroom Activities

Significant differences were found by commissioning source (P <.01), academic class (P = .05) and GPA (P <.01). A significant interaction was also found between source and GPA (P <.005).

Grading Procedures

No significant main effects or interactions were found on this variable.

Discussion

In terms of overall cognitive complexity (Perry Position) as measured by the Learning Environment Preferences Scale, officer candidates who were enrolled in ROTC programs nationwide scored significantly higher than students enrolled in service academies. The mean Perry position for the ROTC sample was 3.45. The mean for the service academy sample was 3.19. These findings would appear to support the 1977 findings of the West Point study group and indicate a need for service academy cadets to have more opportunities and time to investigate problems in which there is ambiguity and multiple points of view.

In the area of degree of relativism, significant differences were found by GPA. This is explained by the differences in group means between those subjects who had the lowest GPAs (0.00-1.99) and the subjects in the upper two GPA groups (3.00-3.49 and 3.50-4.00). This finding would appear to indicate that perhaps higher levels of academic performance in college for officer candidates are related to the adoption of relativistic forms of thought.

With regard to the students' view of learning, the ROTC group was once again significantly higher. This could be indicative of a greater acceptance of their responsibility to actively participate in their learning rather than to adopt an attitude of a "passive receiver" of knowledge. The significant difference by GPA can be explained in terms of differences between the bottom two GPA groups and the upper three groups. It would appear that as GPA rises, so does the students' recognition that learning is an active process in which the end result is more than just the "approved solution".

In terms of the significant differences found in classroom activities, the ROTC sample, once again, was the highest group. This would indicate they selected classroom activities as most significant to them which involved and/or required higher levels of diversity and ambiguity than those selected by the service academy sample. The differences by class indicate that freshmen and seniors are the highest in this area and significantly different from sophomores and juniors. The highest GPA group was, once again,

416

significantly different from the lowest two GPA groups on this variable. This finding would, again, appear to support the contention that higher levels of performance in college are related to the acceptance and expectation of ambiguity and diversity as beinglegitimate in the classroom.

The results of this study clearly indicate that in the areas of overall cognitive complexity, view of learning, and classroom activities the officer candidates enrolled in ROTC programs are significantly higher than the officer candidates enrolled in the service academies sampled. These findings would seem to suggest that little has changed since the report of the West Point Study Group (1977). Situations laced with ambiguity and diversity still appear to be viewed as illegitimate by the service academy subjects. Situations with no clear answers are still more troublesome for the service academy sample than for their ROTC contemporaries. The data would appear to indicate that the freshman year is the one in which cognitive complexity is most seriously arrested at a service academy. It would appear that students at service academies do, in fact, "catch up" with their ROTC contemporaries by the time they graduate. However, it must be noted that the academies recruit the top 10% of all college students who take the scholastic aptitude tests. With this selection group, it would appear to be a "great waste" that these students have to play "catch up" with anyone.

Further research is needed which would measure how students who rank in the top 10% of all those who take the scholastic aptitude tests and attend civilian universities score on assessments of cognitive complexity. This research would provide valuable data on what happens, in terms of cognitive complexity, to those excellent students who don't choose to attend a service academy. It would also provide a more accurate point of comparison for evaluating how well service academies foster cognitive development in their students.

References

Bryant, C. (1982). The developmental design of Behavioral Sciences 330: Using the measure of intellectual development. Unpublished manuscript, University of Maryland, College Park.

Mentkowski, M., Moeser, M., and Strait, M. (1983). Using the Perry scheme of intellectual and ethical development as a college outcome measure: A process and criteria for judging student performance. Vol I and II. Office of Research and Evaluation, Alverno College.

Moore, W. S. (1986). The learning environment preference instrument: An experimental objective Perry measure. Center for Applications of Developmental Instruction.

Phillips, D. J. (1984). Correlates of intellectual development at the United States Military Academy. Unpublished doctoral dissertation, Teachers College, Columbia University.

Rice, J. F. (1985). A descriptive study of moral and intellectual development of studer⁀s at a military academy. Unpublished doctoral dissertation, University of Denver.

Air Force Women:
Accommodation to the Military Social World[1]

Karen O. Dunivin
Northwestern University

Abstract

The social world of the military is masculine as
characterized by its combat mission and androcentric
culture. In interviews with 35 Air Force female
officers, this study describes how women accommodate
to the prevailing masculine social world, assimilating
its structure and culture in order to succeed.
Specifically, women react to the military social world
in two ways: role and ideological accommodation. Such
accommodation explains why women remain and thrive in a
difficult work environment. It also explains why women
feel mainstream (versus marginal) in a predominantly
male world.

One way to examine military women's roles and statuses is with
the sociological concept of "marginal man" theory (Park, 1928;
Stonequist, 1937; Simmel, 1950; Turner, 1964). This theory
contends that certain individuals are caught between two cultures,
thus occupy peripheral social positions in both cultures. For
example, immigrants to a new country do not belong to the old
culture back home, nor are they assimilated into the new culture;
therefore they fall between two worlds and experience marginality.
Military women may be considered marginal persons because of
contradictory social statuses (Hughes, 1945; 1949). Their ascribed
status (sex) conflicts with their achieved status (occupation).
There is a status dilemma and confusion of social identity because
cultural norms define soldiering as a male vocation. Consequently,
women occupy marginal positions in the military.
Previous research on military women (e.g., Goldman, 1973;
Binkin and Bach, 1977; Holm, 1982; Segal and Segal, 1983) examined
servicewomen's marginal position from an external perspective,
focusing on exogenous factors (structural, cultural, and social
barriers) that marginalize women. This study steps beyond the
external realm, analyzing military women's sentiments -- do they
feel marginal; how do they reduce feelings of marginality? While
exogenous barriers may foster marginalization of women, if they do
not feel marginal, they likely do not see themselves as marginal
members in the Air Force or the military.
Data analysis of the research indicates that women experience
and feel some marginality in their roles and statuses as military

women, particularly in gender roles (e.g., woman or mother) which often conflict with occupational roles (e.g., officer or pilot). However, they reduce feelings of marginality in two ways: through role accommodation and ideological accommodation. First, they formulate a compatible role identity in the androcentric social world of the military. Second, they articulate an ideology that helps them view the social world favorably.

Method

The study was conducted in 1987 as part of my doctoral dissertation. I used a questionnaire and an interview guide to collect data on the women's sense of subjective marginality. For example, I asked questions about how they joined the Air Force, their experiences as women and leaders in the Air Force, and their opinions on several issues (e.g., sense of commitment to the Air Force, role demands and conflicts, and servicewomen's current roles and statuses).

Thirty-five women were interviewed at four different bases, representing six major commands and four command levels. The tape-recorded interviews lasted from two to five hours in length and were transcribed for analysis. To ensure response comparability, a questionnaire and an open-ended but focused interview guide were used. However, there was ample flexibility in the interview for free expression and new direction.

This study relied on qualitative research -- a structured, empirical method would have limited analysis of the women's feelings and perceptions. I am aware that the methods of this study and the style of presentation are vulnerable to criticism from colleagues in the military and social sciences. Admittedly, the small, convenience sample makes inferences and generalizations to the larger population of Air Force or military women statistically questionable. The anecdotal presentation raises questions of respondent self-reporting and fair representativeness in the use of data. These inherent problems must be acknowledged. The only answer to these criticisms lies in the quality of the work itself -- in its ability to capture the essence of social reality as described by these women. For these women, their stories describe the social world in which they live and work.

Results

The masculine-dominated military (including the Air Force) forces women to accommodate to its social world. The numerical skewness of women (i.e., 10 percent of total forces) creates a token social world where women must accept its system and culture in order to succeed (Law, 1975; Kanter, 1977; Rustad, 1982).

Military women in this study reacted to the token world of the military in two ways: role accommodation and ideological accommodation -- there were too few women to change the masculine world so they adapted to the social world. First, they formulated a compatible role identity. Then, they articulated an ideology to help them view the Air Force social world favorably.

How did the women formulate a compatible identity? First, they developed a "work self." Women engaged in a process whereby they imagined how they look to others. The "others" in this case were the women's reference group -- members of the officer corps,

420

primarily influential male senior officers. Next, they imagined how others judge their actions. They used the reference group as the evaluator. Finally, they adjusted their behaviors to conform to the others' judgments. In the process, as Cooley (1909) points out in his discussion of the looking-glass self, the reference group became a mirror in which the women compared and evaluated themselves and their behaviors. From their interaction with the reference group, the women constructed a work self.

Through this process, women adopted the men's world and its perpective. From this perspective, it was rational to react by adapting their roles to conform to male standards. Specifically, the women emphasized work roles that are rewarded in the Air Force, and downplayed devalued gender roles. Now it is clear why they behaved as described in this study. For example,

1. The women subordinated personal and family roles to work roles. They sacrificed marital and parental roles in favor of military roles. Or those who chose to combine work and family accepted potential career compromise. They also compartmentalized professional and private lives to reduce role conflict. Indeed, 40 percent of the women delayed or cancelled plans for marriage or children in order to continue their military careers. Three-fourths of the single women (12/16) delayed or cancelled marriage, citing potential career conflict:

> I don't think women officers who aspire for the
> higher ranks can afford to be married...For a man
> to make general he must have a family and a wife.
> For a woman at the higher levels, it's a real handicap.

2. They did not develop gender consciousness -- a sense of identity with women or women's issues. They distanced themselves from other women, focusing on work roles. As a consequence, they did not network (professionally or socially) with other women, and viewed them as competitors:

> Most of the women I run into...view me as competition
> ...there's a lot of animosity. Lots of times I try
> not to be competitive, but other women seem competi-
> tive, so I end up responding that way.

3. They viewed themselves as the classic "exception" to women, adopting an individualistic sense as characterized by Friedan (cited by Freeman, 1975:viii): "there's men, there's women, and there's me." As a result, they evaluated women from the men's perspective (the culture they assimilated), and denigrated fellow Air Force women as incompetent or stereotypical women:

> I find most women are incompetent. I've seen a lot
> more airhead, bubblehead females than males in the
> Air Force.

> Men are more aggressive and they know what they want
> ...A lot of women aren't as aggressive or assertive:
> unsure of what they want.

In addition to role accommodation, women constructed an ideology harmonious in the military social world. As a result,

they perceived a world which allowed them to be upbeat and
optimistic about the roles and statuses of military women,
including themselves. Women of this study displayed three
strategies to construct a compatible ideology, thus reducing
feelings of marginality. They:

     1. Engaged in social comparison with other women to feel
"relative fortune" versus "relative deprivation."

     2. Downplayed the token evironment by de-emphasizing incidents
of sexual harassment.

     3. Evaluated military women's roles and statuses from their
individual perspective versus a broader, collective perspective.

1. Social Comparison

     One way to develop a congruent ideology is to compare yourself
with less fortunate persons, thus feel more successful. For
example, over three-fourths (77 percent) of the women described
themselves as more successful in their careers than civilian female
friends or acquaintances. They cited factors such as better pay,
job challenges, leadership opportunities, and travel as indicators
of their more successful careers.

     Interestingly, women compared themselves to other women when
describing their successes:

> I was promoted three years below-the-zone [early
> promotion] to major. I looked at the statistics,
> and there were only three women promoted three
> years below-the-zone that year. I thought that
> was really good to be one of the top women.

Or they compared servicewomen's present roles and statuses to
former days, again citing improvements.

> We've [women] come a long way. I see it better
> than it used to be, and I'm talking almost 20
> years...women are integrated, maybe not fully,
> but they are integrated.

     Comparing self to less fortunate women or more restrictive
social times, permitted the women to highlight their gains and
accomplishments. Instead of "relative deprivation" they felt
"relative fortune." Had they compared their careers and lives with
men (e.g., job and promotion opportunities or combining work and
family), they probably would have felt some degree of relative
deprivation and marginality. However, to maintain a complementary
ideology in the Air Force, women must focus on positive aspects of
their lives and careers. Negativism is a defeatist attitude that
fosters feelings of marginality.

2. Downplay Hostile Environment

     A second way to maintain a compatible ethos in the Air Force
was to downplay a hostile environment. Women reduced feelings of
marginality (feeling like outcasts) by de-emphasizing negative
experiences as women in a man's world.

     The most obvious example of this attempt to ignore a hostile
climate was their reaction to sexual harassment and discrimination.

Nearly three-fourths of the women described personal experiences of sexual harassment and discrimination (ranging from sexist comments to explicit sexual requests and physical contact). However, none of the women did anything about it. Many were fearful and reluctant to report incidents because of potential negative repercussions. Some women were afraid of receiving bad ratings:

> I had a boss that...openly made sexist comments about women...I ignored him. I was honestly afraid to speak up because of possible repercussions on my OER [officer effectiveness report].

Others were embarrassed and fearful of even discussing the matter. For instance, before describing her sexual harassment experience, one woman said,

> I'm glad this conversation is just between you and me, and let me talk a little quieter because someone could be listening [we were in a private office with the door closed].

Consequently, it was easy to persuade them to take no action:

> [After being slapped on the butt by a colonel], I felt embarrassed and didn't know what to do. I went to another colonel and told him what happened and he said, "Oh, he probably didn't mean to do it." But, I knew he meant it. I was young, scared, and trusted the other colonel's opinion, so I didn't pursue it; I just dropped it.

Such defenses reduced women's feelings of sexist treatment, and helped them ignore harassment as an overt technique to make women feel as outsiders in an organization. Some women rationalized sexual harassment as an occupational hazard in a masculine-dominated organization. For example, one woman reported,

> There is one guy in our squadron who calls all women "bitches." You have to chill out because if I worried about it every time he said it, God, I'd be on the phone to Social Actions all day. I think that is part of it -- women have to adapt to being in a man's Air Force.

3. Individualists

In order to develop a congruent ideology in the Air Force, the women evaluated servicewomen's roles and statuses based upon their own experiences. By focusing on self (a successful woman who beat the odds to be a career officer in a sex-atypical organization), the women ignored issues that affect women as a collective -- sex segregation of occupations, sexual harassment, or lack of career advancement to top leadership positions. Their individualistic approach allowed them to focus on women's successes versus failures. They did not have to think of the thousands of women who had left the military because they had experienced sexism. They could overlook women's denied access to jobs, money and power because they had attained a piece of the pie.

Their individualistic focus also permitted the women to distinguish themselves from other women -- there's men, there's women, and there's me. From this perspective, they felt more mainstream in the Air Force because they were not the stereotyped "typical woman" who was often denigrated as weak or incompetent. Instead, they saw themselves as individualists whose supreme efforts and achievements promoted assimilation into the male world.

Discussion

The majority of women (83 percent) in this study reacted to the token social world of the Air Force in an individualistic and pragmatic way -- adapting their roles and principles to harmonize with the prevailing military social world. Assimilation strategies such as role and ideological accommodation helped the women feel as though they belong (not only integrated in the military structure but assimilated in the military culture).

For the time being, women will continue their assimilation efforts. Their few numbers restrict their ability to change the social world. So they will emphasize work roles and construct a harmonious ethos in the dominant masculine world of the military.

Women who want to remain in the military and develop a career cannot afford to view themselves (or women) as marginal members because that basic premise requires them realistically to evaluate their career opportunities as marginals, and the prospects are dim. As a result, the women do not recognize or challenge even fundamental inequitable policies and rules that marginalize women's roles and statuses: combat restrictions, strength "quotas," or occupational segregation (e.g., few women flyers).

Although this study examines only Air Force female officers, results of this investigation seem worthy of follow-up and expansion. More data need to be collected, especially from enlisted women and servicewomen in other branches (Army, Navy, and Marine Corps). A comparative study of women in the various grades and military services (and their unique social worlds) will distinguish varying levels of objective marginality (barriers) and subjective marginality (feelings as outsiders).

References

Binkin, M., & Bach, S. S. (1977). Women and the military. Washington, D.C.: The Brookings Institute.

Cooley, C. H. (1909). Social organization. New York: Scribner.

Freeman, J. (1975). The politics of women's liberation. New York: David McKay Co., Inc.

Goldman, N. (1973). The changing role of women in the armed forces. American Journal of Sociology, 78, 892-911.

Holm, J. (1982). Women in the military: an unfinished revolution. Novato, CA: Presidio Press.

Hughes, E. C. (1945). Dilemmas and contradictions of status. American Journal of Sociology, 50, 353-359.

Hughes, E. C. (1949). Social change and status protest: an essay on the marginal man. Phylon, 10, 58-65.

Kanter, R. M. (1977). Men and women of the corporation. New York: Basic.

Law, J. L. (1975). The psychology of tokenism. Sex Roles, 1, 51-69.

Park, R. E. (1928). Human migration and the marginal man. American Journal of Sociology, 33, 881-893.

Rustad, M. L. (1982). Women in khaki: the american enlisted woman. New York: Praeger.

Segal, M. W., & Segal, D. R. (1983). Social change and the participation of women in the american military. Research in Social Movements, Conflicts and Change, 5, 235-258.

Simmel, G. (1950). The stranger. Pp. 402-408 in K. H. Wolff (ed.), The sociology of georg simmel. New York: Free Press.

Stonequist, E. V. (1937). The marginal man: a study in personality and culture conflict. New York: Scribner.

Turner, R. (1964). The social context of ambition. San Francisco, CA: Chandler Publishing.

Attitudes Toward Women's Roles in the Military as a
Function of Gender and Race

Michael D. Matthews & Eddie C. Melton
Drury College

Charles N. Weaver
St. Mary's University

Abstract

In the current study, attitudes of a national sample of Americans toward the role of women in the military was examined. Data on 1800 respondents to the 1982 General Social Surveys was used to examine the relationship of gender and race on whether women should or should not occupy eight different military jobs ranging from traditional female jobs such as typist, to nontraditional jobs for women such as fighter pilot. Results indicated that women tended to be more favorable than men toward women serving in nontraditional/combat jobs, but men tended to be more favorable of women serving in traditional or nontraditional/noncombat positions. Blacks were generally less supportive of women serving in any military role than whites, but the interpretation of this is complicated by race covarying with other demographic variables. The results are discussed in context of military and social planning.

Since World War II, women have assumed an increasing number and diversity of roles within the military. A considerable body of research now exists on the capacity of women to adapt to these roles and to perform the jobs expected of them (examples of recent papers on this topic from this conference include Bridges, 1984; Cummings, 1986; Fatkin, 1984; & Savell, 1986). As women enter job specialties formerly held exclusively by men, controversy often follows. Despite an increased awareness of women's issues in the general society, traditional attitudes toward women's roles in society at large remain rather conservative (Deaux, 1984).

Women's participation in "nontraditional" (that is, formerly all-male) jobs in the armed forces also remains controversial. Military concern of a possible negative impact on readiness for war as a function of increasing numbers of women in nontraditional jobs led to the sponsoring of an unpublished report conducted by the Air Force Human Resources Laboratory (1981) on this issue. This report concluded that the gender of workers had little if any impact on wartime mission capabilities.

Despite a large literature in the general domain of sex differences in the military, little objective information exists concerning the attitudes of the general public toward the utilization of women in various military roles. The purpose of the current study was to examine, using a large nationally representative data base, attitudes toward women in three types of military jobs — traditional (eg., typist), nontraditional/noncombat (eg., mechanic), and nontraditional/combat (eg., fighter pilot) as a function of the gender and race of the respondent. Identifying demographic correlates of attitudes toward women in the military may assist in the establishment of plans and policies pertaining to the assignment of women to military occupations.

Method

Sample and Materials Responses from 1800 participants in the 1982 General Social Survey (GSS) conducted by the National Opinion Research Center provided the data base for the current study. The GSS involves face-to-face interviews of a national sample of United States citizens age 18 and over. Respondents are selected randomly, and the sample is stratified to insure representative sampling of major demographic groups. The GSS poses over 300 questions to each respondent on a wide variety of demographic and attitudinal topics.

In the 1982 GSS, respondents were asked to indicate whether a female should or should not be allowed to serve in eight different military roles. These jobs included nurse, typist, commander of a military installation, mechanic, transport pilot, jet fighter pilot, anti-aircraft gunner, and hand-to-hand combat soldier. Two of these jobs are traditionally female (nurse and typist), three are nontraditional/noncombat jobs (mechanic, military commander, and transport pilot), and three are nontraditional/combat (hand-to-hand combat, anti-aircraft gunner, and fighter pilot).

Responses to these questions were crosstabulated with (1) gender of respondent; and (2) race of the respondent (blacks vs. whites only, because other groups had an insufficient N). Estimates of statistical significance were made using Chi Square, and an alpha of .05 was selected.

Results and Discussion

Gender. The responses of males and females to the acceptability of women performing various military jobs are shown in Table 1, which shows the percentage of respondents of each sex who stated that women should or should not serve in each of the eight jobs examined. Several interesting observations may be drawn from the data. First, the distribution of responses of males and females differs significantly for each job considered. Because of the large N, small differences are significant, but they do represent reliable patterns of responding. Concentrating on "should" responses, the greatest differences between the attitudes of males and females involved transport pilot, military commander, and hand-to-hand combat soldier. Overall, there were three jobs for which women held more favorable attitudes: military commander, jet fighter pilot, and hand-to-hand combat soldier. Importantly, each of these jobs may be classified as nontraditional, with two involving direct combat. With the exception of anti-aircraft gunner, all other jobs in which men showed higher approval rates were those that could be classified as traditionally female (typist and nurse) or nontraditional but noncombat (mechanic and transport pilot). It may be that respondents did not view anti-aircraft gunner as a direct combat role in the same sense as fighter pilot or combat soldier, where one-on-one battles may ensue. Thus, women appear somewhat more liberal in their attitudes toward female involvement in combat military roles, while men are more supportive of women's participation in traditional and nontraditional/ noncombat jobs.

Race. Table 2 shows percentage of white and black respondents who indicated that women should or should not occupy each of the eight military jobs. In examining responses on the basis of race, a clear pattern of blacks being less approving of women in any military role is evident. Significant differences

exist for each comparison with only the jobs of typist, military commander, and hand-to-hand combat soldier showing differences of less than five percentage points. Although blacks tend to be less approving than whites in attitudes toward women serving in the military, the rank order of responses for both races are quite similar. Thus, both blacks and whites are most approving of women serving as typists, nurses, and mechanics; and least approving of women serving as hand-to-hand combat soldiers.

It is probably not appropriate to attribute the differences in white and black opinions to race per se. Other evidence from the data base used in this study suggests that variables correlated with race, such as economic level and education, are also strongly related to approval. Thus, before any definitive statements concerning race and attitudes could be made, these variables would need to be controlled.

Davis, Lauby, and Sheatsley (1983) report an analysis of the same data base used in the current study. Although they analyzed different combinations of variables than those used in the current study, their conclusions regarding the attitudes of Americans toward women in the military can also be drawn from the current analysis. First, "Americans are not for total equality of the sexes in the military" (p. 35). That is, participation in direct combat is not viewed favorably by Americans, although as seen in the current study, women are more approving of such roles than are men. A second conclusion drawn by Davis et al. (1983) may also be made here, which is that Americans are in favor of a greater role of women in the military than is perhaps generally recognized. Thus, the majority of Americans favor the participation of women not only in traditional jobs such as nurse or typist, but also in every nontraditional job examined in the current study except hand-to-hand combatant. This conclusion may be modified to point out that some groups such as blacks, older Americans, and the poorer and less educated may be substantially less approving than the population at large.

The data reported in this study and that of Davis et al. (1983) may have implications for military policy makers and recruiting personnel. If decisions to integrate women into a wider variety of military roles are made, military decision makers must expect varying degrees of approval from different segments of the population. Similarly, how women are depicted in recruiting efforts could be influenced by the nature of the target group. It may be acceptable to depict women in a variety of military roles, but showing them in clearly traditional roles or in clearly combat roles may elicit reactance on the part of the target population.

In conclusion, the data from the current study indicate how two demographic variables — gender and race — are related to attitudes toward women participating in a variety of jobs in the military. It is clear that such variables are central to a complete analysis of these attitudes, and need to be considered in coming to a more complete understanding of sex-roles in the military. The results have important implications for military planners and decision makers, and for social scientists interested in women's issues as a whole.

REFERENCES

Air Force Human Resources Laboratory. (1981). The impact of personnel issues on mission effectiveness study: Phase 1: Conus Interviews Report. Unpublished report, Air Force Human Resources Laboratory, Manpower and Personnel Division, Brooks Air Force Base, Texas.

Bridges, A. C. (1984). Changes in attitude toward women at the Air Force Academy. <u>Proceedings of the Ninth Psychology in the Department of Defense Symposium</u>. Colorado Springs, Colorado: United States Air Force Academy, Department of Behavioral Sciences and Leadership.

Cummings, W. C. (1986). Where the men are men: Sex-role perceptions of male and female cadets at the Air Force Academy. <u>Proceedings of the Tenth Psychology in the DoD Symposium</u>. Colorado Springs, Colorado: United States Air Force Academy, Department of Behavioral Sciences.

Davis, J.A., Lauby, J., and Sheatsley, P.B. (1983). <u>Americans view the military</u>. (Technical Report/NORC Rep. No. 131). Chicago, IL: National Opinion Research Center.

Deaux, K. (1984). From individual differences to social categories. <u>American Psychologist</u>, <u>39</u>, 195-216.

Fatkin, L. T. (1984). Stress and military women: The relationship of job and life experience to menstrual distress. <u>Proceedings of the Ninth Psychology in the DoD Symposium</u>. Colorado Springs, Colorado: United States Air Force Adademy, Department of Behavioral Sciences.

Savell, J. M. (1986). Gender-related situational factors in sex-role attitudes measurement. <u>Proceedings of the Tenth Psychology in the DoD Symposium</u>. Colorado Springs, Colorado: United States Air Force Academy, Department of Behavioral Sciences.

Table 1

Gender

Position	% Answering should		% Answering should not	
	Male	Female	Male	Female
Nurse	90.4	90.8	8.5	7.3*
Typist	97.0	95.4	2.6	2.8*
Military Commander	52.3	58.6	45.5	37.3*
Mechanic	80.1	79.6	19.0	17.5*
Transport Pilot	73.0	65.4	25.7	30.6*
Jet Fighter Pilot	57.0	59.8	41.6	36.2*
Anti-aircraft Gunner	57.0	52.1	40.9	43.0*
Hand-to-Hand Combat Soldier	29.6	36.0	69.1	59.5*

* denotes statistical significance

Table 2

Race

Position	% Answering should		% Answering should not	
	White	Black	White	Black
Nurse	93.2	83.8	5.8	13.1*
Typist	96.8	94.1	2.4	3.4*
Military Commander	57.2	53.2	40.3	41.5*
Mechanic	83.6	70.2	15.3	25.3*
Transport Pilot	62.1	49.7	36.0	44.8*
Jet Fighter Pilot	72.3	58.9	25.7	35.8*
Anti-aircraft Gunner	59.3	41.6	38.1	52.1*
Hand-to-Hand Combat Soldier	34.1	31.4	63.5	63.2*

* denotes statistical significance

Women in the Military: The Politics and Women's Roles

Carolyn H. Becraft
Decision Resources Corporation

Abstract

Public policy decisions regarding the role of women in the United States
military since World War II have been made in response to political needs
rather than changing military doctrine. The declining pool of eligible males
for military service plus the lack of career opportunities for mid-career
military women will continue to direct political decisions to increase the
opportunities and numbers of women in the U.S. military well into the next
decade.

One of the more emotional public policy dilemmas for politicians is the role of
women in the military. While insisting that they won't stand for women in combat,
politicians in both political parties have been responsible for the ever expanding
numbers and roles of military women. As the pool of males eligible for military
service continues to decline well into the next decade, there will continue to be
political pressures resulting in increases in opportunities for women in the military.

The increases in the percentage of women in the military and the number of
career fields open to them are the direct result of political and legal decisions in the
past 20 years. *Three specific political decisions paved the way for the dramatic
increase in numbers and job opportunities for women:* the elimination of ceilings on
the grades and numbers of military women in 1967; the end of the male draft resulting
in the All Volunteer Force in 1973; and the opening of the military academies to
women in 1976. Three court decisions were also significant. Two of these decisions,
Frontiero v. Richardson ruling that dependents of military women could receive the
same entitlements as the those offered for the dependents of military men, and
Crawford v. Cushman ruling that a woman could not be discharged if she became
pregnant or if she had dependents under the age of 18, removed significant
impediments to the retention of military women. A third decision, Owens v. Brown
forced the Navy to open categories of ships for assignment to women. As a result of
these decisions, and concerted pressure from Defense officials in the Carter
Administration to expand opportunities for women, the percentage of women in the
military increased dramatically, from less than 2 percent in 1972 to 8 percent in 1980.
By 1987 the percentage of women in the military had risen to 10.2 percent and this
trend is expected to continue.

Currently over 220,000 officers and enlisted women serve in the military, many in
nontraditional "combat type" jobs. Women today are an integral part of the U.S.
military presence and have participated actively in military missions. For example, in
the 1983 Grenada invasion, Air Force women flew planes that delivered supplies and
equipment to U.S. Forces, and Army women, serving in a variety of combat support
roles, were present from the second day of the invasion. In 1986 Navy women pilots
performed carrier landings as a part of the anti-terrorism operation against Libya and
Air Force women pilots were a part of the crews that provided combat support to the

fighter planes that attacked Libya. In 1987 when the U.S.S. ACADIA was sent to the Persian Gulf to repair damage done to the frigate U.S.S. STARK, 25 percent of crew of the U.S.S. ACADIA were women.

There is no U.S. law that prevents women from serving in combat; there are, however, statutory provisions for the Air Force and the Navy restricting the permanent assignment of women from duty in certain places. In the Air Force women, with the exception of medical, dental, chaplain and other "professionals", are prohibited from duty in aircraft engaged in a combat mission. Currently only the offensive fighter aircraft and selected reconnaissance planes are designated as combat aircraft; therefore, approximately 97 percent of Air Force jobs are available for assignment to women.

Navy women are prohibited from permanent assignment on vessels or aircraft that can be expected to be engaged in a combat mission. The responsibility for designation of the particular categories of aircraft or ships as having a "combat" mission has been left up to the Navy. Navy women can be assigned to support aircraft, repair ships, and most recently to certain support ships in the Combat Logistics Force.

The Marine Corps, part of the Department of the Navy, is the most restrictive in its policies for utilization of women. The Marine Corps adheres to the constraints of the Navy and further restricts women from serving in combat units or combat situations as designated by the Marine Corps.

Unlike the Air Force and the Navy, the Secretary of the Army has the statutory authority to determine the assignment policies for all soldiers. In 1977, prior to the final disestablishment of the Women's Army Corps, the Army developed a combat exclusion policy that resulted in the classification of 38 enlisted job categories as "combat" jobs unavailable for assignment to women. In 1981, in a backlash to the pressure from the Carter administration to increase the numbers of women, the Army announced plans to reevaluate the role of women, which culminated in the closing of 23 additional job categories in 1982 (Holm, 1982). Almost immediately the Army experienced criticism on the hypothesis and methodology utilized in its policy review. Critics included representatives from academia, manpower specialists, commanders in the field, advocacy groups, Congress, and the Secretary of Defense's own Defense Advisory Committee on Women in the Services (DACOWITS). In 1985, after a review directed by the Secretary of the Army, many of the job categories closed in 1982 were reopened. Currently 49 job categories are closed to women.

As noted earlier, the various services are not consistent in their policies concerning the utilization of women. Women can be pilots in all of the services except the Marine Corps. Civilian women serve regularly on civilian ships that deploy with the Naval battlegroup, while Navy women are more restricted in assignments to parallel Navy ships in the Combat Logistics Force. All positions in the Coast Guard have been open to women since 1978, yet in time of war Coast Guard ships will come under the control of the Navy which limits opportunities for women at sea. Army women are present in the forward support areas of the battlefield while Marine Corps women could conceivably be banned from the continent.

The original "combat exclusion" laws were part of the Women's Armed Services Act of 1948 (P.L. 80-625) which was passed to provide a means of mobilizing women in the event of general war. Defense Department officials and prominent military leaders strongly favored the inclusion of military women in the Regular forces and did not favor combat exclusion provisions. However, Congressional opponents, in an effort to ensure that women would not be employed as combatants, would not agree to passage

of the Act without exclusions reflecting the political views of the time (Treadwell, 1954; Holm, 1982).

In the ensuing decades, the influence of technology has resulted in weaponry, some with intercontinental ranges, that dramatically enlarge the battle area. Military strategy and doctrine have changed to utilize the advanced technology. Current doctrine calls for deep first strikes into the rear of the battle area to knock out the supply and service base of the opposing force, thereby crippling the combat arm. It is acknowledged that it is unlikely that there will be a stationary front line (Department of Army, 1982). In short, in the past, military battles resembled a football game with opposing forces fighting head to head to gain precious yardage. Future military scenarios will be more like soccer, with a front line constantly on the move over large pieces of territory, deep attacks into the rear, and a constant reliance on the defense.

Military doctrine has changed to reflect the realities of the 1980s technology, but technology has yet to influence assignments and opportunities for women. Rather, opportunities for women have expanded and the numbers of women have increased because of the political need to rely on a volunteer force, the increased willingness of women to enter nontraditional occupational fields, and the inability of the services to recruit adequate numbers of qualified men. The definition of "combat" has changed frequently when defining women's assignments and positions, but the changes have been rooted either in the service's recruiting needs or in politics without regard to doctrine.

Members of Congress, politicians, and defense advisory groups are once again pushing for expanding opportunities for military women. Because of increasing recruiting difficulties as a result of the declining pool of eligible males for military service, attention is being focused on the potential need for increased numbers of women in each service. In addition, the varying interpretations of the combat exclusion laws and policies by the services have resulted in inconsistencies in the types of positions available for women as well as a shortage of career opportunities for qualified mid-career women. In an effort to ease these conditions legislation was introduced in the House and the Senate in 1987 would require the services to open thousands of combat support jobs in all services to women.

Public support is another aspect of the political dimension in the expansion of roles for military women. Arguments against increasing opportunities and roles for women in the military invariably begin with concerns about the "will of the American people." Evidence suggests, however, that the public is quite comfortable with nontraditional roles for military women. A poll conducted in August of 1986 by NBC-TV News found that 52 percent of the public supported women in combat support roles and 77 percent were comfortable with the fact that military women would not be evacuated in a military conflict. This coincides with a 1982 poll by the National Opinion Research Center which measured support for military women in specific jobs. According to these data, 84 percent of the American public supported keeping or increasing the proportion of women in the services. The approval rate for military women in combat support jobs was equally high: 83 percent approved of women as military truck mechanics; and 72.7 percent as jet transport pilots. In the more offensive military jobs 62 percent reported support for women as missile gunners, 57 percent as crew members on a combat ship and 35 percent as participants in hand to hand combat (Davis, Lauby, & Sheatsley, 1982).

The issue of the role of military women will continue to be on the political agenda. The declining pool of eligible male recruits, coupled with a Congressional and Department of Defense desire to maintain the quality of recruits that the services

433

realized in recent years, and the insufficient variety of career opportunities for qualified mid-career women will continue to force increased scrutiny on the services interpretation of combat exclusion laws (Binkin, 1984). The alternative to enlarging the pool of eligibles by increasing opportunities for women would be to return to the all male draft to make up for recruiting shortfalls, a politically unpalatable solution. Therefore, the U.S. military will continue to need the quality and the skills that the women provide. So will the politicians.

References

Binkin, M. (1984). <u>America's Volunteer Military: Progress and Prospects</u>. Washington, D.C.: The Brookings Institution.

Crawford v. Cushman, 531 F. 2d 114 (2d Cir., 1976).

Davis, J., Lauby, J., & Sheatsley, P.B. (1982). <u>Americans View the Military: Public Opinion in 1982</u>. University of Chicago: National Opinion Research Center.

Department of Army. (1982). <u>Operations, FM 100-5</u>. Washington, D.C.: Author.

Frontiero v. Richardson, 411 U.S. 677 (1973).

Holm, J. (1982). <u>Women in the Military: An Unfinished Revolution</u>. Novato, CA: Presidio Press.

Owens v. Brown, 455 F. Supp. 291 (D.D.C. 1978).

Treadwell, M. (1954). <u>The Women's Army Corps</u>. Washington, D.C.: Department of Army, Office of the Chief of Military History.

Enhancing the Quality of Life for Military Spouses:
The Issue of Underemployment [1]

Jeannette R. Ickovics, M.A.
Walter Reed Army Institute of Research

Abstract

This article reviews the current literature on the
nature and effects of the employment situation for
military spouses. Military wives have lower rates of
employment, lower earnings, and higher rates of
unemployment and underemployment compared to their
civilian counterparts. An emphasis will be made on
identifying some of the causes and consequences of
underemployment in this population. Relevant policy
issues will also be addressed.

The most striking change in the labor force since World War
II has been the dramatic influx of women. There has been a 200%
increase in the number of women in the labor force during this
40-year period, compared to a 50% increase in the number of men
during the same period (Bureau of National Affairs, 1986). With
this change has come increased interest and concern about the
health effects of employment for women. It was expected that
women's morbidity and mortality rates would increase to resemble
those of working men. To the contrary, a wide variety of studies
that compare the physical and mental health of women who are
employed versus those who are not, find that employed women have
a decided advantage (cf, Baruch, Biener & Barnett, 1987 for a
review).

In the military, spouse employment has been linked not only
to individual health, but to the health of the military
institution -- namely through readiness and retention. In a
recent study completed by the Civilian Personnel Center
(CIVPERCEN, 1985), 62% of soldiers reported that family member
employment is important to retention decisions (reported by Army
Research Institute, 1987). Work, in addition to being an
economic necessity, can be a source of identity, and it enables
many military spouses to establish their connection to society as
valuable and contributing members. This commitment to spouse
employment has challenged the military institution in its
readiness and retention efforts. It has also had an impact on
the provision of child care services, employee placement and
assistance, relocation and other family policy initiatives.

----

[1] NOTICE: The views of the author do not purport to reflect the
position of the Department of the Army or the Department of
Defense, (para 4-3, AR 360-5). Special thanks to LTC James A.
Martin for his insightful contributions to this manuscript.

Military Spouse Employment

Consistent with the rapid increase of labor force participation among civilian wives, military wives have also been joining the work force in increasing numbers. According to the Bureau of Labor Statistics (1984), labor force participation among military wives increased from 36.3% to 51.1% in the decade from 1974 to 1984. Many of these women appear to have entered the labor force for reasons of personal growth and satisfaction, as well as economic necessity (Martin, 1985).

Unfortunately, while many military wives may want to and need to work, they are at a decided disadvantage in securing meaningful employment. Due to their lifestyle, they are confronted with a number of obstacles which impede their efforts to secure employment commensurate with their education and skills. As a result, military wives have lower rates of employment, lower earnings, and higher rates of underemployment and unemployment compared to their civilian counterparts (Grossman, 1981).

Unemployment

Military wives have consistently had a harder time finding employment than other married women since at least 1970, when the Bureau of Labor Statistics began collecting data on military wives. The unemployment rate of military wives is currently more than three times the unemployment rate of their civilian counterparts. The unemployment rate for civilian wives improved slightly between 1981 and 1984, decreasing from 5.6% to 5.4%. Among military wives, however, the unemployment rate more than doubled during the same period -- rising from 8.0% to 17.0% (Bureau of Labor Statistics, 1984).

While the reasons for this dramatic increase are not clear, many factors can be identified which contribute to the overall high rate of unemployment among military wives. Most detrimental to their employment status are the frequent moves from one military installation to another. Military families are relocated an average of once every three years. These constant moves preclude opportunities for job stability, the development of seniority and job advancement. Many military wives find that they must start at entry-level jobs each time they move, regardless of their former employment experience. Moreover they are often faced with discrimination because of their military connection; some classified ads read "locals only" and some employers have denied job interviews to those women whom they consider to be short-term workers (Grossman, 1981). Constant moves also reduce the chance for completion of occupational and educational training programs. The difficulty of getting a job is further compounded by the fact that many military installations are located overseas or in remote geographic areas where job opportunities may be scarce.

These barriers often result in military wives accepting employment outside of their fields of expertise, particularly

women who are in professional or technical fields (Burgess, 1984). In fact, the lack of adequate employment opportunities for spouses was cited as the most serious problem facing enlisted service members in a PCS move -- regardless of assignment location. Finding employment for their spouses was listed as a problem by 41% of enlisted members assigned in the continental U.S. as well as 48% of members assigned overseas (Young, 1986).

Underemployment

Unemployment is a visible problem which continues to receive substantial attention from both social scientists and policy makers. The unemployment rate is regularly measured and reported in monthly labor statistics; it has become an important economic indicator and frequently a source of political inquiry. However, the emphasis on labor force participation and unemployment may mean that other problems of labor utilization are neglected (Sullivan, 1978). Inadequate employment is one such problem. Underemployment tends to be "invisible" by virtue of the economic structure and official labor statistics. People who are underemployed do work, but they are subject to intermittent employment, involuntary part-time work, overeducation, marginal jobs, being workers in poverty households and inadequate compensation (United States Commission on Civil Rights, 1982).

Persons in the "employed" category when calculating labor force statistics are categorized as such regardless of the quality or quantity of work the employed person performs. By acknowledging the heterogeneity of employed persons, and moreover, by creating a subcategory of those who are underemployed, we can begin to understand which jobs or job characteristics are beneficial and which are harmful. While even unskilled employment can provide social stimulation and feelings of competence (Belle, 1982; Ferree, 1976), viewing one's work as a career rather than as a job is associated with greater satisfaction and less role conflict. Contrary to the belief that the more high-powered a woman's career, the more dangerous it is to her well-being, the advantages of employment are greater for women in occupations with higher status (Haynes & Fienleib, 1982; Verbrugge, 1987). Jobs that combine high demands with low autonomy are most likely to have negative health consequences; these jobs are typically low-status and low-paying ones.

Jacques (1976) claims that every employee has the right to abundant employment and individual opportunity. By this he means "the right of every individual who takes up employment...to be able to find work reasonably in line with his interests and experience at a level consistent with his capabilities" (p. 182). Jacques hypothesizes that the consequences of underemployment are boredom, monotony, restlessness, resentment and despair. Family members, too, may be adversely affected by the restricted employment opportunities which limit family income and increase financial pressures. Further, the underutilization of household members may also result in anxiety, depression, distress and decreased well-being (Foster, 1981; Kelly, 1979).

438

Given that many miltiary spouses face constant underemployment, the potential effects of chronic underemployment for communities should also be considered.

> Long periods of under-employment -- or indeed a career of underemployment -- can lead to a chronic semi-depressed resignation in people and a lack of conscious awareness of their true capacity. Somewhere inside something tells the individual that he is capable of greater things, but he hardly believes it.... Without the opportunity [to realize his capabilities] the individual may continue in his chronic state of torpor. Whole communities can be forced into such a state by chronic underemployment or unemployment (Jacques, 1976, p. 184).

Jacques (1976) suggests that these symptoms can only be alleviated by giving the individual the opportunity to make use of his or her capabilities.

Policy Implications

The Department of Defense has been addressing the employment needs of service member spouses by incorporating policies such as preferential hiring in federal jobs. However, this does not assist those spouses who are employed or who seek to be employed in civilian jobs. The services should also consider policies which are aimed at reducing discrimination against military spouse employment in both the government and the private sector.

Policies which promote better employment opportunities, rather than just jobs, must also be explored. It is vital to enhance the quality of women's employment experiences in order to enhance the health and well-being of military spouses, families and the institution in general. For instance, job training for military spouses should be directed toward higher-level and higher-paying jobs, and toward jobs with skills which are transferable from location to location. Entrepreneurship in "portable businesses" or cottage industries may also be a productive outlet for highly motivated military spouses. Issues relating to child care and after-school care must continue to receive attention given the prevalence of dual-earner families. Finally, because constant mobility is most detrimental to employment opportunities for military spouses, extending tours of duty in order to increase family and job stability must also be considered.

Equality in employment is central to the battle against sex discrimination. Work quality as well as its availability is a relevant issue. Research should be designed to understand employment experiences and the consequences for military spouses and their families across a career. Moreover, in order to best address these issues, it is important to be sensitive to differences in race and ethnicity, social class and rank. Policies and programs alone, however, will not be effective

439

unless they are accompanied by changes in values and expectations about both service member and spouse participation in work and in family roles.

References

Army Research Institute. (1987). Report presented at the Army Family Research Program In-Progress Review, October 27, 1987. Community and Family Support Center: Arlington, VA.

Baruch, G., Biener, L. & Barnett, R. (1987). Women and gender in research on work and family stress. *American Psychologist*, 42(2), 130-136.

Belle, D., (Ed.) (1982). *Lives in Stress: Women and Depression*. Beverly Hills, CA: Sage.

Bureau of Labor Statistics. (1984). Labor Force Status of Military Wives and Civilian Wives, March 1970-1984 (Chart). Bureau of Labor Statistics, Washington, DC.

Bureau of National Affairs. (1986). *Work and Family: A Changing Dynamic*. Bureau of National Affairs: Washington, DC.

Burgess, T. (1984, November 5). Three military wives jobless for every civilian one in U.S. *Army Times*, pp. 3, 66.

Ferree, M. (1976). The confused American housewife. *Psychology Today*, 10, 76-80.

Foster, A. (1981). Wives earnings as a factor in family net worth accumulation. *Monthly Labor Review*, 104, 53-57.

Grossman, A. (1981). The employment situation for military wives. *Monthly Labor Review*, 60-64.

Haynes, S. and Finleib, M. (1982). Women, work and coronary heart disease: Results from the Framingham 10-year follow-up study. In P. Berman and E. Ramey (Eds.), *Women: A Developmental Perspective*. (Publication No. 82-2298, pp 79-104). Bethesda, MD: National Institutes of Health.

Jacques, E. (1976). *A General Theory of Bureaucracy*. London: Heinemann Press.

Kelley, J. (1979, August 27). Pay causing Navy's retention woes. *Navy Times*, 1.

Martin, J.A. (1985). The wives of COHORT soldiers: Initial information from a longitudinal study of the Army's New Manning System. *Medical Bulletin of the U.S. Army*, 42, 14-18.

Sullivan, T. (1978). *Marginal Workers, Marginal Jobs*. Austin, TX: University of Texas Press.

United States Commission on Civil Rights. (1982, November). *Unemployment and Underemployment among Blacks, Hispanics and Women*. Washington, DC. Clearinghouse Publication #74.

Verbrugge, L. (1987). Role responsibilities, role burdens and physical health. In F. Crosby (Ed.) *Spouse, Parent, Worker*. New Haven, CN: Yale University Press.

Young, S. (1986, July 21). U.S., overseas families share same problems. *Army Times*. p. 50.

Employment Among Air Force Spouses
at Los Angeles Air Force Base

Lieutenant Colonel Victor N. Ogilvie, Ph.D.
Los Angeles AFB, California

Abstract

A growing national trend toward full-time employment of military spouses has raised important local questions about possible effects on recruiting and retention of scientific and technical officers at Los Angeles Air Force Base. A locally developed and scored survey was administered to spouses of USAF members and to key supervisors at Los Angeles AFB. Results indicated that spouse employment would have some influence on an Air Force member's willingness to accept PCS assignment orders (in lieu of retirement/separation) outside the Los Angeles area, but would not be a major obstacle. Prior spouse employment did not significantly lessen the USAF member's willingness to relocate to accept their current Los Angeles assignment. Somewhat paradoxically, spouse employment was strongly supported by senior officers at Los Angeles AFB with the degree of importance attached to a spouse's career actually increasing in direct proportion to the grade of the officers being surveyed.

While our country's military has usually paid salaries to the individual service members, total financial support to the families has not always been fully adequate. As a result, many wives, especially those of low-ranking enlisted members, have been forced to supplement their family income through outside part-time or full-time employment (Hunter et al., 1981; Himaka, 1977). During the last half of the twentieth century, the number of wives who work for reasons other than financial need has dramatically increased. A Ladycom survey (1977) of 9,600 active duty wives of all ranks revealed that many wives work because they desire to, as opposed to having to do so. At Los Angeles AFB it has been estimated that as many as 64% of the military spouses work outside the home. Yet, despite this increasing trend, some commanders are not always supportive and may exert pressure on the military member for a spouse to curtail paid employment in lieu of base-level volunteer work (Fisher, 1987).

Lewis et al. (1986) in a study of 811 couples (USAF enlisted men and their wives) discovered that wives who work outside the home tend to have more positive attitudes toward the Air Force than their non-working counterparts. Wives who had a positive career orientation (i.e., think it important for a wife to have her own career), particularly those not currently employed, have more positive attitudes toward the Air Force than those who don't value working. Moreover, those with a positive career orientation were more involved in the Air Force lifestyle.

Because wives have traditionally assumed the major responsibility for

household chores and child care, it seems logical that these demands would be more difficult for a working wife to satisfy than one who is unemployed and able to be a full-time homemaker and actively support her husband's military career, especially by participation in such organizations as OWC, NCO Wives' Club, Red Cross, Family Services, Little League, etc. One must wonder what will be the future of these and similar organizations if more wives will be entering the workforce. Another question that must be addressed is the possible effect of spousal employment on permanent change of station (PCS) into and out from Los Angeles AFB. A spouse's reluctance to abandon or suspend a career could decrease a husband's willingness to move to Los Angeles or to leave the area for another assignment. A third, but related issue, is whether a spouse's employment is viewed as a career liability for the Air Force member by Commanders and other key supervisors. The appearance of less than total support of the Air Force and it's various family programs, other responsibilities notwithstanding, could conceivably bias a supervisor, even if all other aspects of an Air Force member's performance and bearing are otherwise unblemished.

Method

Subjects

Male and female spouses of officers and enlisted Air Force members were surveyed regarding their attitudes toward employment. Since 832 married Air Force members are serviced by the Los Angeles AFB CBPO, a 50% random sampling was used by taking every other name on an alphabetical listing of all assigned personnel. One hundred eighty two (182) completed questionnaires were received. Key supervisors were defined as officers in grades O-4 through O-6. Therefore all 317 field grade officers at Los Angeles Air Force Base were surveyed regarding their attitudes about spouse employment. A total of 210 completed questionnaires was received.

Instruments

Two questionnaires were constructed to measure: 1) spouses' attitudes toward and involvement in USAF activities, and attitudes toward PCS waves, and 2) attitudes of key supervisors regarding spousal employment.

Procedures

Mailing labels were obtained from the Los Angeles AFB Consolidated Base Personnel Office and the surveys were mailed to the persons previously identified. Instructions at the top of each survey provided full directions on completing and returning the forms. A letter from the Base Commander explained the purpose and importance of the survey and assured complete anonymity for all respondents. Since the surveys were both voluntary and anonymous, no follow-up was possible for those persons declining to participate. Completed surveys were collected by the base Family Support Center.

Results

Part I- Military Spouse Employment Survey Questions

1. Are you personally <u>interested</u> in Air Force activities for spouses (e.g. wives clubs, Red Cross, Youth Center, Family Services, etc.)?

63 Yes
90 No
23 Indifferent

Of those who answered yes, 38 narrative responses were received and the majority indicated strong interest in either on-going base-wide activities (e.g., youth center, family services, Red Cross and chapel activities) or in wiíː/spouse groups, organizations or activities. Minor scattered interest was received in several unrelated activities.

Expressions of negative interest in Air Force activities were received by 100 individuals. Reasons most oftenly cited involved lack of time due to job, educational programs, or childrearing responsibilities. A second major reason was excessive geographical distance from Los Angeles Air Force Base or Fort MacArthur. Other miscellaneous reasons given for lack of interest were involvement in other (i.e. non-USAF) activities, desire for privacy, and lack of activities for male spouses.

2. Do you <u>participate</u> (past or present) in Air Force activities for spouses at Los Angeles AFB or Fort MacArthur?

35 Yes
145 No

Of the 35 spouses who indicated that they did participate in activities for spouses, 28 provided written responses. Wife/spouse groups, organizations or activities were the most popular outlets of involvements with 20 individuals participating. A small minority of persons also participated in various on-going base-wide activities (e.g. base picnic, etc.). Of the 145 spouses not participating, 121 explained why. Lack of time due to employment, school and/or childrearing was the primary reason given by 45 (37.2%) respondents. Another 39 (32.2%) individuals gave geographical distance as the main obstacle. Personal reasons - 16 (13.2%); lack of awareness of activities - 13 (10.7%); and general lack of interest - 8 (6.6%) were the other reasons given for non-participation.

3. If you are <u>presently employed</u> and your military spouse received PCS assignment orders outside the Los Angeles area, how much would your job or career affect his/her ability or willingness to relocate and accept that assignment? (Please check one).

13 Would not accept the assignment (separate/retire from AF)
 9 Might not accept the assignment
14 Unsure, cannot determine
22 Would probably accept the assignment

51 Would definitely accept the assignment
                              or
    59 I am not employed at present

    4. If you were previously <u>employed</u> <u>before</u> <u>moving</u> to the Los Angeles
area, how much did your employment affect your military spouse's eventual
willingness to relocate here? (Please check one).

        12 I was very reluctant to leave my job
        11 I was somewhat reluctant to leave my job
        64 Not a factor in deciding to move
        13 I was somewhat eager to leave my job
         4 I was very eager to leave my job
                          or
        59 I was not working before coming to Los Angeles

    Part II - Key Supervisors Survey Questions

    1. Do you think it is important for Air Force wives to have careers of
their own (if they so desire)?

	Total	0-6	0-5	0-4	Grade ?
Very important	130	28	50	46	6
Important	69	7	26	30	5
Indifferent/no opinion	13	1	4	8	0
Unimportant	5	1	1	3	0
Very unimportant	4	1	0	2	1

    2. If one of your subordinate managers had a spouse who was employed
full-time, outside the home, would you evaluate him/her lower (on a OER)
<u>only</u> because of the inability of the spouse to attend certain Air Force
functions due to her being employed?

	Total	0-6	0-5	0-4	Grade ?
Definitely would rate lower	0	0	0	0	0
Probably would rate lower	0	0	0	0	0
Indifferent/don't know	2	0	2	0	0
Probably would <u>not</u> rate lower	22	3	11	8	0
Definitely would <u>not</u> rate lower	186	34	70	80	2

    3. Would you be less willing to give this officer/manager certain high-
level duties or assignments <u>only</u> because of the spouse's inability to attend
certain Air Force functions due to her job duties?

	Total	0-6	0-5	0-4	Grade?
Definitely would be less willing	0	0	0	0	0
Probably would be less willing	5	1	3	1	0
Indifferent/don't know	2	0	2	0	0
Probably would <u>not</u> be less willing	18	3	9	6	0

Definitely would <u>not</u> be less willing   184   33   68   81   2

Discussion

It does appear that the trend toward greater spouse employment will decrease participating in traditional Air Force social and avocational activities at Los Angeles Air Force Base. Nearly two thirds of local Air Force spouses are employed, and the majority of them are working full time. While there is moderate interest in these activities, the lack of available time (because of employment and childrearing responsibilities) leaves very few opportunities for actual participation by interested spouses.

Spousal employment would have some influence on an Air Force member's willingness to accept PCS assignment orders outside the Los Angeles area, however this influence is limited and only a small minority of Air Force members would actually separate or retire in order to remain in the local area and to preserve their spouse's career at the expense of their own. Similarly, prior spouse employment did not appear to significantly lessen the USAF member's willingness to relocate to the Los Angeles area in order to accept their current Space Division assignment.

A working spouse cannot be considered a career liability for the Air Force member at Los Angeles Air Force Base. Spouse employment is strongly supported by senior officers at LAAFB. Indeed, the degree of importance attached to a spouse's career actually increases in direct proportion to the grade of the officers being surveyed, with Colonels being more supportive than Lieutenant Colonels and then Majors, in that order. Senior officers showed no inclination to give lower OERs solely because of a working spouse's inability to attend certain Air Force social functions, and no reluctance to give military subordinates increased responsibilities only because of a working spouse's inability to attend USAF social functions because of civilian job duties.

References

Fisher, J. (1987, November 23). Officers' wives feel heat to stay in kitchen. <u>Orlando Sentinel</u>, pp A-1, A-7)

Hunter, E.J., Gelb, B.M., & Hickman, R.A. (1981). <u>Military wife adjustment: An independent dependent</u>. (Report No. TR-USIU-81-05). Family Research Center, United States International University.

Ladycom survey results. October 1977). <u>LADYCOM</u>, pp 22-4; 36; 52.

Lewis, P., Winther, D., Adan, A., and Dansby, M. R. (1986). Changing marital roles and the support of Air Force wives. <u>Proceedings of the Tenth Psychology in the Department of Defense Symposium</u>, 649-653.

Pierce, M. M. and Luchsinger, M. L. (1986). Psychological distress among Air Force wives. <u>Proceedings of the Tenth Psychology in the Department of Defense Symposium</u>, 643-648.

Dual-military couples: A descriptive study

Major Nancy K. Raiha, ACSW, Ph.D
9th Infantry Division (MTZ)
Fort Lewis, WA

Abstract

A growing segment of the military population is comprised of
dual-military couples, in which both husband and wife are active duty
military members. This study analyzes the responses to a written
survey concerning demographic characteristics, morale-related issues,
and retentions plans, which was completed by a probability sample of
20,000 active duty Army members. The responses of the 1350 soldiers
who identified themselves as married to another service member were
compared to responses of other married and single soldiers. The
resultant profile of dual-service couples and their concerns provide
implications for policy, practice, and future research.

The changes in military family life prompted by the rapidly increasing
number of two service-member marriages have evoked much speculation and
comment (Finegan, 1983; Hunter, 1982; Landrum, 1979; Maze, 1987; Orthner,
1980; "Wedlock, Deadlock," 1982). Observers have noted that current family
policy has not changed to reflect new family structures and women's changing
societal role (Bowen, 1985; Hunter, 1981; Kohen, 1984; Martin, 1984; McCubbin,
Marsden, Durning, & Hunter, 1978). The failure to adequately address
two-soldier family needs may have a negative effect on mission accomplishment,
soldier performance, retention, and personnel policies (Akerlund, 1986;
Margiotta, 1980). Better knowledge of the dual-career family is needed to
guide policy and practice efforts (General Accounting Office, 1982).

Method

In February 1983, the Soldiers Survey Division, Soldier Support Center,
National Capital Region, distributed a written survey to a systematic sample
of all active duty Army personnel. Respondents were chosen by terminal digits
in the social security number so that approximately 10% of all officers and 5%
of all enlisted members would be surveyed. The more than 20,000 responses
represent nearly 50% of all eligible officers and enlisted soldiers. More
than 1350 respondents identified their spouse as another active duty service
member. This author (1) created a profile of the dual-career soldier
encompassing demographic, morale-related, and retention-related factors; and
(2) contrasted the characteristics of these dual-career soldiers with
characteristics of other married and single soldiers.

Results

More than six percent of the survey respondents were involved in
dual-service marriages.[1] Of the married respondents, 7.6% of officers and
12.1% of enlisted soldiers were married to another service member. Over 60%
of married female and less than 7% of married male soldiers reported
dual-service marriages. Dual-career soldiers tended to be younger and lower

ranking, with less time in service and fewer children than other married soldiers. When compared to single soldiers, dual-career soldiers were older, higher ranking, more experienced, and better educated. Members of dual-service couples were more likely than other soldiers to live off-post, to be involuntarily separated from family, and to have combat service support professions than other soldiers. Within the dual-career group, males tended to be older, higher ranking, and more experienced than females. In the enlisted ranks, the majority of dual-career males were black, while the majority of dual-career females were caucasian.

In general, dual-careerists reported more job satisfaction than single soldiers, but less than other married soldiers. Dual-careerists with children reported more job satisfaction than childless dual-careerists. Dual-career respondents were less likely than other married soldiers to voice support for a strong military, to support universal military service, to feel that today's soldiers are better off than ever before, and to willingly recommend military service to friends and relatives. Dual-career married respondents tended to reflect a greater awareness of the problems of sexual harassment and discrimination than other male and married female soldiers. Dual-career enlisted soldiers had a lower perception of unit climate than other soldiers; they had fewer positive feelings about co-workers than other married soldiers. Dual-career and non-dual-career married soldiers did not differ in overall experience of military-related family/marital problems. However, dual-careerists tended to report such problems later in their careers, while other married soldiers experienced more problems during the first term.

Dual-career males were more likely to plan retention than their single counterparts, but less likely than other married males. Dual-career women's relative commitment to military service appeared to increase with rank. Within the dual-career group, males were more likely to plan retention than females. In the enlisted sample, when demographic variables and job satisfaction were controlled, dual-careerists had lower retention plans than any other group. Dual-career respondents, especially women, were more likely than other groups to choose pay as a reason to remain in the service, but more often chose separation from spouse as a reason to leave the service. Dual-careerists were less concerned with retirement benefits than other married respondents.

Discussion

Many of the demographic characteristics of soldiers involved in dual-military marriages reflect cultural tendencies, the nature of military service, and the relatively high proportion of women in this military population subgroup. The tendency for dual-careerists to be younger and lower ranking than other married soldiers reflects both the recent large-scale acceptance of women in the military workforce and the cultural tendency for women to marry men of similar, but slightly higher, age and status. Women tend to cluster in combat service support fields, and to date and marry men in these fields. The tendency for couples pursuing two demanding careers to have fewer children than more traditional couples has been documented in both the civilian and military literature (Bowen, 1985; Gilbert, 1985).

This large-scale representative sample survey shows that dual-career soldiers are somewhat less satisfied and less likely to enthusiastically espouse military service than soldiers married to civilians, but more satisfied than single soldiers. Findings about more specific morale-related issues tend to support findings of military researchers who conducted in-depth

interviews or family-oriented surveys. This study showed that although dual-careerists were no more likely to report military-related family problems than other married soldiers, they were likely to experience such problems later in their careers. Orthner (Finegan, 1983, p. 8) explains a similar finding: "They were older, more of them were having children, and more of them were being stressed by separate assignments." A number of studies (Farkas & Durning, 1982; Teplitzky, Thomas, & Nogami, 1986; Tice, 1986; Williams, 1978; Wood & Flores, 1982) highlight childcare, separation, and career progression as major issues for the dual-career service member. The dual-career concern with sexual harassment and discrimination noted in this survey may highlight one aspect of the concern with female career progression. Equal opportunity has been found to be a major concern of women considering a military career (Adde, 1986; Hill, 1983; Maze, 1984; Nichols, 1985; Waite & Berryman, 1985). The stresses experienced by dual-military families may be exacerbated by the social isolation reflected by the dual-career respondents to this survey who had a relatively low perception of unit climate and few positive feelings about co-workers. The busy schedules of dual-military couples and the lack of a partner to perform the traditional wife's function of establishing and maintaining social contacts may well contribute to this social isolation. Dual-career soldiers with children may be more satisfied than other dual-career soldiers because of attrition: Only the most satisfied dual-careerists choose to face the logistic problems of remaining in the service after childbirth.

The retention plans outlined by survey respondents appear to reflect security needs, available options, and previously discussed family-related concerns. Traditional "provider" status appeared to increase retention plans; parents and married males, especially those married to civilians, tended to plan retention. A second family income may afford dual-careerists the option of enduring longer periods of unemployment or education in seeking new employment. This greater freedom from security demands, coupled with the inconveniences of a dual-military lifestyle, probably contributed to the trend for dual-careerists to be the least likely group to plan retention when job satisfaction and demographic variables were held constant. The dual-career respondents reported that pay and benefits contributed positively to retention plans, but that possible family separation was a major concern, especially for women. The tendency for dual-career women to have lower retention rates than dual-career men and, in most cases, single women has been documented by other researchers (Orthner, 1980; Teplitzky, Thomas, & Nogami, 1986). Probable reasons include childcare concerns, a tendency for enlisted women to view military service as an escape from home rather than a career (Maze, 1984a), and fewer years invested in military service than the husband. Throughout the sample population, involuntary family separation--more common in dual-career families--was associated with drug use, difficulty finding worthwhile friends in the unit, negative attitude toward military service and benefits, military-related family/marital problems, and intention to leave the service.

In conclusion, dual-military couples can be a productive, useful, and potentially cost-effective element of the Armed Forces (Department of the Air Force, 1979; GAO, 1982; Orthner & Bowen, 1980). These couples and military agencies, however, express concerns about separation, childcare, deployment, maternity/paternity leave, discrimination, and career progression. Unless these concerns are carefully addressed by military researchers, policymakers, and mental health and social service providers, these couples may become a liability--or a loss--to the military system.

References

Adde, N. (1986, May 19). At the crossroads: The first female academy graduates are now making career decisions. Army Times, pp. 80-81.

Akerlund, K. (1986, May 19). Working couples: Who's the boss. Army Times, p. 86.

Bowen, G. L. (1985). Families in blue: Insights from Air Force families. Social Casework, 66, 459-466.

Department of the Air Force (1979). Final report of the workshop on single member parents and military couples with dependent children. Washington, D. C.: Author.

Farkas, A. J., & Durning, K. (1982). Characteristics and needs of Navy families: Policy implications (DTIC #ADA120807). San Diego: U.S. Navy Personnel Research and Development Center.

Finegan, J. (1983, February 7). Service marriages: A booming trend poses problems for the military and the couples. The Times Magazine (Supplement to Army Times), pp. 4-9.

General Accounting Office. (1982). Army needs better data to develop policies for sole and inservice parents (DTIC #ADA119396). Washington, D. C.: Author.

Gilbert, L. A. (1985). Men in dual-career families: Current realities and future prospects. Hillsboro, New Jersey: Lawrence Erlbaum Associates.

Hill, T. M. (1983). Facing issues caused by the increasing female presence in the military: Reflections on a workshop venture. Military Chaplain's Review: Womens' Issues (Department of the Army Pamphlet 165-136, Vol 12, No. 1). Washington, D. C.: Department of the Army.

Hunter, E. J. (1981). Family roles in transition in a changing military (DTIC # ADA107348). San Diego: United States International University, Family Research Center.

Hunter, E. J. (1982). Families under the flag: A review of military family literature. New York: Praeger.

Kohen, J. A. (1984). The military career is a family affair. Journal of Family Issues, 5, 401-418.

Landrum, C. S. (1979, June). The conflicts surrounding family and children versus mission responsibilities. (DTIC #ADA113373) Monterey: Naval Postgraduate School.

Margiotta, F. D. (1980). Evolving strategic realities: Implications for U.S. policy makers. Fort McNair, Washington, D. C.: National Defense University Press.

Martin, J. A. (1984). The military wife. Medical Bulletin of the U. S. Army, Europe, 41(1), 12-18.

Maze, R. (1984, December 3). Moskos urges chain of command for women. Army Times, p. 16.

Maze, R. (1987, May 4). DOD told to be fair to dual-service couples. Army Times, p. 51.

McCubbin, H. I., Marsden, M. A., Durning, K. P., & Hunter, E. J. (1978, September/December). Family policy in the Armed Forces: An assessment. Air University Review, pp. 46-57.

Nichols, C. (1985, June 3). An officer and a gentlewoman. The Times Magazine (Supplement to Army Times), pp. 4-13.

Orthner, D. K. (1980). Families in blue: A study of married and single parents of families in the U. S. Air Force (USAF Contract F3360-79-C-0423). Greensboro, North Carolina: Family Research and Analysis.

Orthner, D. K., & Bowen, G. L. (1980). Families in blue: Implications of a study of married and single parent families in the U. S. Air Force (USAF Contract F3360-79-C-0423). Greensboro, North Carolina: Family Research and Analysis.

Raiha, N. (1987). Dual-career couples in the U.S. Army: A descriptive study. Dissertation Abstracts International, 71, 4513A. (University Microfilms No. 87-06, 654).

Teplitzky, M. L., Thomas, S., & Nogami, G. Y. (1986). Dual Army career officers (Working Paper from Technical Utilization Area, Army Research Institute). Washington, D.C.: Army Research Institute.

Tice, J. (1986, June 30). Researchers track careers of young officers. Army Times, p. 3.

Waite, L. J., & Berryman, S. E. (1985). Women in nontraditional occupations: Choice and turnover. Santa Monica, CA: Rand Corporation.

Wedlock deadlock. (1982, November 29). Army Times. p. 21

Williams, J. W., Jr. (1978). Dual-career military families. In E. J. Hunter (Ed.), Military Families: Adaptation to change (pp. 103-110). New York: Praeger.

Wood, F., & Flores, L. (1982). Dual-career family negotiations among U.S.A.F. Academy cadets. Paper presented at the Eighth Biennial Psychology in the Department of Defense Symposium. Colorado: U.S. Air Force Academy, 1982.

Footnotes

[1]Detailed tables and references are available from the author or in Raiha's 1987 work. All between-group comparisons are significant at $p < .05$.

Adjustment to Separation and Reunion Among Wives of Army Husbands Deployed to the Sinai: An Ethnographic Approach

Suzanne Wood
Katharine S. Gravino

Abstract

In the course of conducting a wider anthropological study of army wives patterns of coping and adjusting to life at an Army post, a sample of wives was followed over a nine month period that included the deployment of their husbands on a six-month peacekeeping mission to the Sinai. Communications with the absent husbands, facilitated by the availability of telecommunications with the soldiers in the Sinai, emerged as important to the wives' sense of adjustment to the deployment. Other factors included quality of the marriage prior to separation, family ties and friendships, family health, and community involvement.

This ethnographic study attempts to understand, describe and interpret the adjustment of wives of enlisted soldiers deployed to the Sinai on a six-month peacekeeping mission. A sample of wives at a West Coast post was followed through three waves of interviews -- pre-deployment, mid-deployment, pre-reunion -- and will be interviewed again in a fourth wave, post reunion (December 1987-January 1988). By mid-deployment, communications with husbands emerged as having high salience for all the wives. Many of the women were surprised by the high frequency of communication with their husbands. We decided to pursue this particular issue in more depth and attempt to capture major themes and commonalities among the wives experiences in communicating*. We are, of course, aware that there are many other factors which will affect the women s adjustment to separation and reunion. Thus, data were also collected on other aspects of the women's lives. (Unfinished project; reporting only mid-point and pre-reunion data.)

Method

Two researchers trained in conducting nonstructured interviews and in participant-observation techniques first had open-ended "conversations" with the wives, covering personal biography (parental family, marital history, children, health, religion), feelings about the Army, preparation of children for

*"Waiting Wives: Suddenly We Have So Much to Say". Paper presented at the October 1987 Inter-University Seminar on Armed Forces and Society.

fathers' departure, plans and expectations for the deployment, and participation in formal/informal support systems. New sets of questions were added with each wave of interviews as circumstances changed. For example, at the three month point of the separation we asked "How's it going?" questions, including how the women saw the separation affecting their relationship with husbands; and their frequency of writing and telephoning, and other exchanges. During the last two months of the separation, we focused on changes in communication patterns, changes the women felt in themselves and in their husbands as a result of the separation, and the women's reunion expectations. The current wave of the study (December 1987, January 1988) is designed to elicit the women's feelings about the reunion and the effect in general of the entire deployment experience upon their lives.

After each interview, detailed fieldnotes were immediately generated and coded for age, race, rank, number/age of children, education, length of marriage, wife's work, finances, connectedness with formal/informal support systems, and family communications, among other variables.

In an attempt to observe and absorb something of the general social fabric of the community of task force wives, the researchers attended many meetings, unit functions, private get-togethers, and the departure and arrival of the soldiers. Additional interviews were conducted with chaplains, the Rear Detachment Commander, the unit commander's wife, and other individuals involved with the women during the deployment. Copious fieldnotes were taken.

Subjects

Forty three wives of enlisted soldiers were randomly selected from the deployed unit and represented all stages of the family life cycle found in the unit.

Results

Mid-Deployment

For some wives, intense communication with the absent husbands seemed to ease loneliness, led to fewer "problems"(as defined by the Rear Detachment Commander), made the time go faster, and generally helped the women to feel that the deployment was "not as bad as imagined",as one wife put it. Access between the women and their husbands in the Sinai was made easy by the availability of commercial direct-dial facilities. A TACSET radio telephone system was also available to wives and husbands but was less popular than the commercial telephone because of the inconvenience and lack of privacy. Most women wrote daily and the husbands responded 2-3 times a week on average.

A major theme among the wives at the mid-point of the deployment was that the separation had, in a counterintuitive way, brought husband and wife closer together. For many, this was the first long separation since basic training, and the six month deployment emerged as a period of relative calm, stability and increased marital connectedness. For some couples, communications during this period reached a higher level than when the men were physically at home. For example, several women commented that with boring duty and little else to occupy their husbands' free time, the men in the Sinai concentrated more on their families. This contrasts with the shorter field training separations at home, when little telephoning and letter-writing occurs. These field training separations are seen by wives as major irritants and not condusive to close marital and family relationships.

Telephoning and writing, for some, also served to help sort out shaky relations or give a young couple an opportunity to think about each other, and to grow individually too. Communicating so frequently also played a major role in keeping all couples in touch, not just those with shaky marriages. It enabled them to sort out problems, involve the children with fathers, share physical and emotional loneliness and, in most cases, draw families closer together. The wives felt that keeping husbands informed so frequently of everyday events meant that the husbands were intimately involved with life at home and consulted in household decision making. As one wife said, "He knows everything that is going on back here, like he was here. Nothing will seem strange when he comes back because he will know it all." The main negative aspect of direct dial telephoning, however, was the extraordinarily high cost. Monthly phone bills ranged from about $100 to over $800.

The quality of a marital relationship, and level of communication within a marriage, may influence a family's ability to cope with separation and reunion. At the mid-point of the deployment, we speculated that the unexpectedly high level of communication between families and Sinai soldiers considerably increased couples' ability to "talk" and thus enhanced marital relationships. Although some of the wives did develop independence and assumed greater responsibility for family decision-making, most of the women in our sample appeared to be allowing their husbands to continue to participate in the family from afar. Problems were often discussed and solved jointly on the phone. We contended that, as these communications appeared to ease the adjustment of a wife to separation, so they were likely to ease the wife's adjustment to her husband's return. The volume and quality of contact, particularly by telephone, were likely to temper idealized expectations of reunion with awareness of daily realities. Preliminary analysis of post reunion interviews indicate, indeed, that couples who reported open communications during the separation (and who communicated well in their marriage prior to the deployment) seemed to weather the stresses of reunion relatively easily.

Pre-Reunion

As the time of reunion neared, changes occured in patterns of communication. While a few of the wives loyally continued to write every day, most slowed down. There was a sense of emotional exhaustion among the wives and some exasperation about how time had suddenly slowed down. This was tempered with feelings of triumph and pride at having survived the experience, and excitement and relief about the imminent return.

Reunion had been a major theme in all the telephoning and letter-writing. Many of the wives (and husbands also, as we determined from them in the post-reunion phase of interviews) tended to romanticize that reunion. They built up fantasies and idealistic expectations, which were sometimes fed by the increasingly passionate phone calls and letters that the wives were receiving and sending. Each woman interviewed in the pre-reunion phase had a scenario for the time of reunion and the short-term period after that. She was concerned with preparing the house, preparing herself to handle the husband's relationship with children, preparing for sexual reacquaintance and a degree of "romance". There was a commonly felt nervousness ("butterflies") about the actual moment of meeting, how they and their husbands would act and react initially, an expressed vulnerability as the women evaluated the changes in their lives since parting.

Most of the women were aware of personal changes that had occurred within themselves over the past six months. These had to do mainly with a sense of increased independence. Having reached this pre-reunion point was seen as success by the women. "We have survived!" said one wife. However, while taking pride and satisfaction in having handled the day-to-day problems alone (with some telephonic aid from the Sinai, of course), they were looking forward to having a husband around to take back some of the responsibility. They acknowledged the potential for marital tension in this transfer of family duties. Through family newsletters and army sponsored reunion seminars before the men returned, women were made intellectually aware of possible reunion problems.

As reunion neared, the women were anticipating the return of their husbands and withdrawing somewhat from the close female friendships that had nourished them during the deployment. Wives were unclear whether, or how, friendship patterns would survive after the husbands returned. Moreover, parental families, who had provided emotional and practical support during the deployment, were becoming less important now.

The wives also anticipated that the personal freedoms permitted by the deployment would be curtailed on their husbands' return. The "single" life, the outings with girlfriends, the spontaneous shopping sprees, and dinners in restaurants would all decrease and family routines would re-establish themselves.

In general a "festival" atmosphere developed among the wives as they built up to a state of euphoria about the reunion. A few discussed the possibility of problems in reestablishing relationships and routines. However, the women for the most part were concentrating on the very moment of reunion and were not dwelling too seriously on anything beyond that point.

Discussion

No single factor, such as communicating, can be solely responsible for successful or unsuccessful adjustment to the separation and reunion experience. Rather, we should look - perhaps keeping communication as a general context - at the interrelationships of many factors at work. We have already noted the importance of family ties, friends (both Army and in the community), church, the demands and consolations of children, and the support services of the Army structure as stabilizing factors. Finally, there are those elusive qualities of good judgement, humor, maturity, and personality in general, which play a role in healthy adjustment.

We have already developed a simple coding scheme covering 12 variables which, at the pre-reunion stage, allowed us to rank our sample wives according to likelihood of adjusting well to the reunion. These variables are: whether a marriage appears strong; whether ties with parental families are strong; whether children have been a support or a problem; whether the women have developed good friendships during the deployment; whether they have had a job or "kept busy" with projects; whether there has been a major crisis during the deployment; whether the wife and children have had good physical and mental health during the separation; whether finances are under control; whether the woman has a bouyant personality; whether she has experienced long separations before; and whether she was pregnant during the separation. We shall analyze the post reunion wave of data on these same variables to rank the women again and to look at what variables correlated with a "successful" reunion. In addition to developing this typology for "success", we shall develop a series of case studies illustrating some of the wives' stories.

Typology of Army Families: Coping Styles of Successful, Career Army
Families[1,2,3]

Russell C. Smith, Ph.D.
HQ USAREUR & 7th Army

Abstract

This is a qualitative study of the coping styles of 18 healthy
Army families. Healthy families are defined as those who are
without overt clinical symptoms of pathology, who enjoy military
life, and who have chosen the military as a career. Data were
gathered by means of intensive, unstructured interviews, and by
administration of Moos's Family Environment Scale. Five
distinctive healthy coping styles were identified and are perceived
to be consonant with the lifestyle of military families.

The active duty Army has approximately 400,000 families, roughly one
and a half million people who, on a daily basis, interact with the largest
military system in the world. An all-pervasive culture unto itself, the
Army effects the lives of each one of these people, to a greater or lesser
degree, every day of their lives. This research (Smith, 1987) was begun in
order to look at the effects which this lifestyle has and how individuals
and families structure their environment to cope with it.

Method

The first step in the research was the identification of families to
study. The question was: "How can we tell what is a healthy family?" From
a strict scientific standpoint, there is no agreed upon definition of
health. However, based on prior research and clinical experience, a measure
of health was assumed from an absence of overt pathology. Families who
appeared to be healthy, who said they were happy, who were successful in the
Army system and demonstrated a degree of satisfaction with Army life, and
who volunteered for this research were assumed to meet the needs of this
definition. A variety of methods were used to recruit families.
Recommendations were made by family "experts", commanders, and by the
families themselves. All families were volunteers.

A pilot study, with eight families, was conducted. These interviews
were loosely structured around general questions about Army life. This led
to wide discussions about Army life. In addition, Moos and Moos's (1981)
Family Environment Scale (FES) was administered to each family.

The FES was used an an objective measure to assess the families' environments. For Moos and Moos, the formula of behavior is $B=f(P,E)$, where B is behavior, a Function (f) of People (P) and the Environment (E). The FES consists of ninety true/false questions. The FES is divided into three main dimensions which are divided into ten scales. There are the three relationship dimensions of cohesion, expressiveness, and conflict; the five personal growth dimensions of independence, achievement orientation, intellectual-cultural orientation, active recreational orientation, and moral-religious emphasis; and the two system maintenance dimensions of organization and control.

For the final section of the research, it was decided that a more systematic investigation of the problems and solutions of Army life was needed. A structured survey was developed from the themes uncovered during the pilot study. It was felt that a structured survey would lead to more complete and more easily comparable data. Ten families were selected for the final interviews.

Normative data were collected on the families through specific questions. The families were then asked the general question: "What have you seen as major problems in being an Army family?" After some discussion, they were told: "Stop, let us discuss one of these problems in detail." As the discussion around the identified problem continued, the families were asked why the issue was a problem; the number of times they had encountered the problem; how they had dealt with the problem; what help they had received, and from whom; and the resources that they found available. For each identified problem, their coping strategies, time resources, and types of successes were constantly looked for, discussed, and written down. In the same fashion, more identified problems were discussed. Specific questions were available to use if the families got stuck. A typical question was: "Tell me about how you moved to this military community."

After the discussion of problems in being an Army family, the families were asked to make a list of generic problems. They were asked to list general problems that Army families face and the assistance that Army families have available. After this point, the families were given the FES.

The general purpose of the interviews was to distill the behaviors of the families that were their mechanisms of coping with the military system. While a structured format was used for the new interviews, the responses were open-ended.

Eighteen families, combined from the pilot and final studies, are not representative of the entire Army in terms of surface variables, and this small sample size should not be considered representative of the current total population of Army families. It is, in fact, a skewed selection - but skewed for the purpose of illustrating the point of the research. In general, successful, older career families were interviewed. However, these families had been stationed all over the world. They were once recruits too. It was hoped that this sample was representative of successful career families in the Army.

Results

Two problems surfaced during this research. One problem was the definition of a healthy family. Despite the breath of our operational definition, our definition was not totally satisfactory. Despite meeting our criteria for being a healthy family, some of these families would not meet our stereotypic view of tne "happy" family nor would we choose them for

"best friends". In several cases, our healthy families were neither likable nor happy. Meeting our criteria did not necessarily mean that a family was generically healthy or that they would be so in a different environment. It did say, however, that each family functioned well and had gained satisfaction from their lifestyles within their specific environment.

A second problem with this research was how to categorize these families in analyzing the parameters of healthy families? Two lines of research proved fruitful.

Stinnett and DeFrain's (1985) work with families identified six areas of healthy families. These were: Commitment, Appreciation, Communication, Time, Spiritual Wellness, and Coping Ability. These six characteristics of strong families are an important theoretical construct which gave us clues about healthy Army families. Because the interviewed families, for the most part, exhibited these characteristics, the research focused on the last characteristic, our research variable of coping.

However, analyzing specific coping behaviors was difficult. Part of the problem, and confusion, with this research was the different attitudes of families toward Army life. There were unseen variables operating in their choices. For example, some families preferred to live in Army housing, others wanted to live off-post, some families liked or accepted mandatory social events, while others didn't and/or wouldn't. These wide ranges of likes and dislikes seemed to be too random to make sense.

It did make sense, however, to look at how the families interacted with the Army environment and how they structured their environment. Moos and Moos's (1976) cluster analysis of family FES profiles was used to categorize the families. In their original sample, Moos and Moos identified six major cluster types. These were sub-divided into 12 cluster types of family social environments. Grouping the families under Moos's clusters supplied some order to what appeared to be whimsical preferences and attitudes. While the Army families were originally grouped by their FES profiles, they could be grouped by what they said in their interviews.

Five distinct coping styles were identified from the sample families. These coping styles are from a combination of Moos's 12 clusters types. Each type coped somewhat differently; they viewed and related to the system differently; and, they formed a sense of community differently. It was very interesting to note the differences in percentages of families in the cluster types between Moos's sample and our sample of Army families. While this is not a random sample of Army families, and no claim is made that families in this research are statistically representative of Army families, the differences in percentages of cluster types is striking. In summary, these differences are listed in Table 1.

Table 1

Comparison of Family Cluster Type Distributions

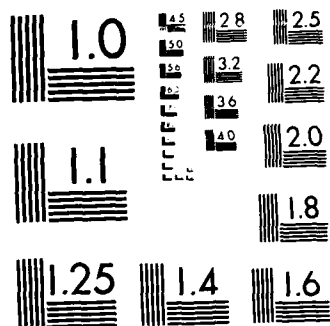|	Moos %	Army %
Apathetic-independence	3	22
Individual personal growth	51	17
Expression-oriented	9	17
Expressive conflict	3	11
Structure-oriented	8	33
Conflict-oriented	26	0
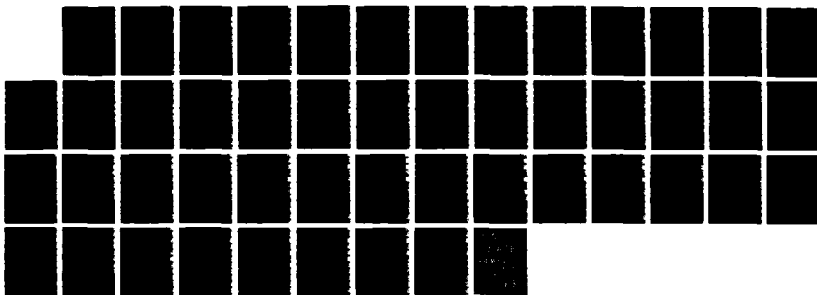
Army families are significantly under-represented in the individual personal growth cluster type and over-represented in the apathetic-independence and structure-oriented cluster types. The sample, when compared to Moos's, is probably skewed by age. Army families, by their nature, are young. This would lead to differences in the family life-cycle stages; but, this is not enough to explain all the differences. The Army families' average FES subscale scores, in general, are well within one standard deviation above or below the mean. This little difference, in a broad sense, seems to indicate that the sample families are fairly representative of families in general. This might indicate that these differences exist, not because of a skewed sample of Army families, but because Army families use specific coping styles.

Apathetic-independence families see the Army as an amorphous structure, and they have a very narrow view of their place in the system. This was our second largest group, and their coping style was very recognizable.

Individual personal growth families include the six separate personal growth dimension clusters of expressive-independence, structured-independence, achievement-via-independence, achievement-via-conformity, unstructured moral-religious, and structured moral-religious. Because these sub-groups are more alike than different, they have been combined into one group. These sub-groups believe in an Army system, but they feel they are separate from it. They believe they can gain command of it through their own personal development. What separates these families from each other are their views of the Army system and their techniques for personal development.

Expression-oriented families view the Army as a means to an end, whereas they appear to have less control over their lives and therefore let the system manipulate them at will. For them, their situation is temporary and therefore managable, as it is a means to their end goal of Army life.

Expressive-conflict families believe the Army can provide things but will probably have to be forced to do its job. Their conflict with system, and each other, occurs within a reasonable structured and controlled environment and appears to produce an adaptation to the system for them.

Structure-oriented families view the Army as an end goal. In other words, they Army provides them with the things they want. They view themselves in and of the structure of the Army. In general, these were the "oldest" families, and they might be the "ideal" end product of almost 20 years of socialization in the Army system.

It was not surprising that there were not any families from the structured and unstructured conflict-oriented family clusters. These families most likely would not meet our definition of a healthy family. They probably would not be successful in the military in attaining much rank or career status nor would they volunteer for this study.

Discussion

We have asked: "What can the Army do to help families?" There seems to be no easy, clearcut answer to this question if families view and deal with the Army system differently. Their coping styles are dependent on how they view the Army system. For example, some types see an amorphous structure while others see a system they have to learn and master. It reasons that Army programs trying to meet the multitude of needs on which such views are predicated are de facto required to be "all things to all people", and thus doomed to failure unless we have a clearer understanding of what, precisely, families mean when they say "I need".

With these five groups of families, we have a spectrum of people's needs and attitudes, with the apathetic-independence families on one end. In general, these families want the Army to provide almost all of their needs, from providing friends to preserving social mores. To best serve them, the Army needs to provide complete, easy programs for them, and indeed, a variety of existing Army programs fit these families.

The other groups of families are on the other end of the spectrum. In general, these families want to be provided with what they consider their just rewards and then left alone. What separates these groups from each other is their involvement with the Army system.

The individual personal growth dimension families feel that they have to work for what they want. They gain, not so much from the structure of the system, but through their individual efforts. They have to feel that they can master the system. The Army needs to provide programs aimed towards personal development that these families can use to grow. The off-duty education program is one such program that the Army provides.

The expression-oriented families feel a lack of control over the system, but also feel the system will take care of them. They get involved mostly with their small work-based units, as they are most comfortable at that level. In general, these are junior and mid-level families, and this cluster type might be the first stage or a passing stage for them as they spend more time in the Army and begin making a more reasoned choice about whether or not to pursue the military as a career. To best serve them, the Army needs to encourage and provide resources for small unit cohesion. These families need to feel they have some control over their small unit activities, as they generally feel that they do not have much control over the formal Army system.

The expressive conflict-oriented families are somewhat different. They feel that they have to fight the system for what they want. It is hard to provide any special considerations for these families, as they tend to want to fight the system, regardless of the ease of access.

Finally, the structure-oriented families feel that they can influence the system. They not only become intimately involved with the system, they need to feel that they belong to it, because it provides the basic structuring of their lives. The Army needs to allow these families to work for the system, because they make the best organizational workers the Army can get. These are the "joiners".

References

Moos, R. & Moos, B. Family Environment Scale Manual. Palo Alto, CA: Consulting Psychologists Press, 1981.

Moos, R. & Moos, B. A Typology of Family Social Environments. Family Process, 1976, 15 (4), 357-372.

Smith, R. C. Typology of Army Families: Coping Styles of Successful, Career Army (Doctoral dissertation, Northwestern University, 1987). Dissertation Abstracts International. (Citation in press).

Stinnett, N. & DeFrain, J. Secrets of Strong Families. Boston: Little, Brown and Company, 1985.

Drill Sergeants' Families: Benign Neglect as Policy?[1]

Pearl Katz, Ph.D.
Walter Reed Army Institute of Research
University of Maryland School of Medicine
The Johns Hopkins School of Medicine
Uniformed Services University of the Health Sciences

Abstract

An ethnographic study of Drill Sergeants' families by an anthropologist at the Department of Military Psychiatry explores the complex interrelationships of soldiers' and families' adjustments to PCS, Drill Sergeant School and Drill Sergeants' duties. This paper describes some initial stresses experienced by families of Drill Sergeants, the effect of the Drill Sergeants' chronic fatigue upon the families, and some ways in which both unit and family stresses interact. It examines the families' utilization of social networks and their consequences for adaptation.

Studies on Drill Sergeants have emphasized their stressful work conditions, that their work dissatisfaction was a major stressor, and that they experienced a large degree of marital and family difficulties (Vernon, 1986; Carney, 1984) The family memorandum issued by Secretary of Defense Weinberger acknowledged that "family problems often have an adverse effect on readiness, retention, and quality performance." (Military Family, 1986:1)

Drill Sergeant duty is among the most stressful work in the Army (Fullerton, 1984). The "Raupp Report" (Task Force on Initial Entry Training Leadership, 1978) and subsequent studies on Drill Sergeants by Carney (1984), Fullerton (1984) and Vernon (1986) found that Drill Sergeants' work stresses negatively affected their morale, work satisfaction, sense of well-being and family life. Vernon's study found that the sources of stress in Drill Sergeants' work roles, such as long hours, few breaks between training cycles and inconsistencies between policy and practice, created difficulties

---

[1]The views of the author do not purport to reflect the position of the Department of the Army or the Department of Defense.

Drill Sergeants' family life. She found that many of the Drill Sergeants reported fragility, friction, and difficulties in their marriages, and that they felt alienated from their families. Vernon concluded that social support among Drill Sergeants enhanced the relationship between Army satisfaction and well-being in a very significant way (1986:19).

Vernon's conclusion supports the large body of literature indicating that social networks play significant roles in buffering and ameliorating the effects medical, psychological, marital, family and work stress. Social networks of family members of soldiers have been found to play a crucial role in soldiers' well-being, their perceptions of stress, and in their marital relationships (Blochberger 1970).

The present study of Drill Sergeants' families focuses upon the processes of social network recruitment and utilization by the soldier and by the spouse. It investigates the roles of the social networks of Drills Sergeants and their families in obtaining information, facilitating access to resources and in facilitating adaptation, and ameliorating stress, as well as the roles of social networks in hindering adaptation and increasing stress.

This research studies the complex interrelationships of soldiers' and families' adjustments to PCS, Drill Sergeant School and Drill Sergeants duties by examining the natural history of the soldier and family during a two-year tour. By focusing upon the interactions between family and Army, it challenges the perspective of viewing Army life and family life as occupying distinct and competing domains.

Method

Subjects

The subjects consist of 30 Drill Sergeants and 28 Drill Sergeants' spouses (2 are single parents) at one Basic Training post. Eight of the Drill Sergeants are female. There are two Drill Sergeant marital couples. They represent a range of ethnic backgrounds. The subjects were selected during the first phase of Drill Sergeant School, soon after their PCS to the training post. They volunteered to be studied throughout their two-year Drill Sergeant tour.

Procedures

This is an ethnographic study which examines the natural history of Drill Sergeants' families as they adapt their work and family life to the demands of Drill Sergeant duties. Two methods are used in this study, open-ended interviewing and participant-observation. In most cases, spouses were interviewed and observed separately from the Drill Sergeants.

Field research began in June 1987, and is presently continuing. The data presented is, therefore, restricted to the early stages of Drill Sergeant families' adjustments. The subjects were initially interviewed shortly after the soldier began studying at Drill Sergeant School. The duration of the initial interview ranged from 4-6 hours. Subsequent interviews ranged from 2-4 hours. Participant-observation took place with Drill Sergeants and their spouses in the course of their daily lives. Drill Sergeants were observed while on duty, and to a lesser extent, at home. Spouses were observed at home and/or at work.

The initial interviews focused upon a) personal, military and family history, b) perceptions of Army life and of marital and family life, and perceptions of the interactions of these two domains, and c) historical background on adaptations to stress and the structure and role of social networks. Subsequent interviews focus on details of their present lives, such as a) daily schedules, including sleep schedules, b) network interactions on-post and off-post, c) communication with spouses and with children, and knowledge of and empathy with spouses' present life, and d) perceptions of stress, adaptation, Drill Sergeants' duties and family life.

Results

The results indicated a number of sources of initial stress on the Drill Sergeants families. Some of the major stressors affected the Drill Sergeant and *families before beginning Drill Sergeant duty.* Many resulted from lack of information about conditions on the post and requirements of Drill Sergeant School. They included a) living in crowded temporary housing from 2 to 6 months, b) being locked-in Drill Sergeant School for three weeks, completely isolated from family; c) having orders for a specific Drill Sergeant class but assigned classes 3 months later (i.e. 1/3 of one class arrived to attend scheduled previous classes), d) having to spend $400 to $600 on uniforms, and having separate rations deducted from their pay.

Most of the married Drill Sergeants arrived with their families expecting permanent housing; many had heard that Drill Sergeants received housing priority. They lived in small, single room temporary quarters from two to six months, and were required to move from room to room each week. The living arrangements provided no privacy, affecting marital relations, study habits of children, and adolescent privacy concerns. The spouses cooked in one common kitchen with one stove. They waited in long lines with other families to cook. These conditions adversely affected the family mealtime atmosphere as well as the kinds of food they cooked. Families reported they developed poor nutritional practices, children suffered from constipation, and spouses and children consumed large amounts of junk food. An additional source of stress resulted from weekly moves to another room in the temporary quarters; families moved all their belongings to different rooms, often on different floors, with the possibility of being told there was no room available.

Drill Sergeant School was stressful for both the families and for the Drill Sergeants. Most received no prior information about the Drill Sergeant School. Specifically, they did not expect to spend $400 to $600 on uniforms, nor to have meal rations deducted during one pay period (in addition to the daily expense of temporary housing). They did not expect to be separated from their families for a three week lock-in: "The day he signed in, he found out he'd be 3 weeks in. He called me to say that he's not coming home for three weeks. The kids don't understand it." In Drill Sergeant School, the soldiers did not expect to be treated like privates. They did not expect that they would sleep an average of 3.5 hours a night.

The greatest stressors for the soldiers, spouses and children were the long hours of work, frequent CQ duty assignments, short or non-existent cycle breaks, and the unpredictability of their schedules. The data indicate that all the Drill Sergeants experienced chronic fatigue, and that all the spouses reported that the Drill Sergeant came home very tired, and fell asleep shortly after arriving home.

The spouses reported drastically reduced marital and family communication opportunities. Communication focused almost exclusively on the most important practical matters. The Drill Sergeant rarely saw their children, and their interaction with their children was cursory. Sexual relationships virtually ceased or, if they occurred were drastically reduced, and unsatisfactory for the spouses.

Each family was affected by the chronic fatigue of the Drill Sergeant. They missed the wakeful, alert, active member who had previously actively participated to a greater extent in family life. However, the families of soldiers who had served in Combat Arms MOSs reported less change, disruption and negative effects upon their family life. Female Drill Sergeants with young children reported greater stress upon their children which was associated with their absence.

Some spouses, however, developed strategies for ameliorating the stressful effects of Drill Sergeant Duty. These include a) Spouses who created social networks which replicated the structure of supportive social networks in the past; b) Spouses who had jobs which provided them with gratification; and c) Spouses who were particularly knowledgeable about the work life of the Drill Sergeant, particularly those who had served, or presently served, in the Army.

In addition, preliminary data suggested that a) some Drill Sergeants' family stresses were buffered or ameliorated by support of the leadership at the Company level, that appeared to reflect good leadership at the Battalion level; and conversely, b) some Drill Sergeants' family stresses were affected by, and believed by them to be caused or exacerbated by, poor Company organization and lack of support of the leadership at the Company level, that appeared to reflect leadership at the Battalion level.

Discussion

The initial results indicate that the stresses upon Drill Sergeants have a profound effect upon their family lives: Each of the Drill Sergeants reported not only less time with his/her spouse and family, but also decrements in the quality of interaction. Some of the causes of family stress, such as the conditions of temporary housing and the lack of information during Drill Sergeant School, appears to be a result of a benign neglect of family life of Drill Sergeants.

Chronic fatigue is responsible for much of the negative effects upon Drill Sergeants' family lives. Drill Sergeant shortages exist in most companies and result in long working hours. However, companies differ in their arrangements for work and CQ schedules, "details," and duties during weekends and cycle breaks. Drill Sergeants' morale, and subsequently many of their families' morale, is also affected by their perception of their First Sergeant's and, to a lesser extent, by their Company Commander's understanding, support and appreciation of their work.

All of the Drill Sergeants reported satisfaction received from training soldiers. However, this satisfaction was rarely shared by their families whose satisfaction they perceived as depending more upon their own individual strategies for maintaining their family lives.

References

Blochberger, C.W. Jr. (1970) Military Families: Differential Lifestyles. Unpublished Ph.D. Dissertation. University Microfilms, 71-9754. Ann Arbor, Michigan.

Carney, C. M. (1984) Well being and distress among drill sergeants: Civilian and military comparisons and the role of social support. In, Proceedings: Psychology in the Department of Defense. Ninth Symposium, 18-20 April, 1984, 148-152.

Fullerton, T. D. (1984) Comparing well-being and stress of several high risk Army groups. In, Proceedings: Psychology in the Department of Defense. Ninth Symposium, 18-20 April, 1984, 143-152.

Military Family (1986). SecDef underscores importance of family advocacy programs. Military Family 5, 1.

Task Force on Initial Entry Training Leadership (1978) Toward Positive Leadership for Initial Entry Training. Fort Monroe, VA: U.S. Army Training and Doctrine Command.

Vernon, D. (1986) The Impact of Social Support and Coping on the Relationship Between Work Satisfaction and Well-Being Among Army Non-Commissioned Officers. Unpublished Ph.D. Dissertation. University of Pittsburgh.

The Identification of Major Dimensions of
Individual Soldier and Unit Readiness: Initial Findings[1]

Melvin J. Kimmel, Ph.D.
U.S. Army Research Institute

Robert Sadacca, Ph.D.
Mary M. Kralj, Ph.D.
Susan M. Heon
Human Resources Research Organization

Abstract

The U.S. Army Research Institute for the Behavioral and Social Sciences recently established the Army Family Research Program. A major goal of this program is to determine the linkages between family factors and Army community services and individual soldier and unit readiness. A prerequisite for establishing these relationships is the development of instruments that adequately measure readiness. This paper describes the procedures that were used to develop initial sets of rating scales for assessing individual soldier and unit readiness. The paper presents the dimensions identified to date and the results of evaluations of the initial dimensions by Army officers and NCOs.

The recently published Army Family Action Plan III clearly states the goal of fostering wholesome lives for Army families. The plan further states, "This objective follows the realization that by providing for families, the Army is contributing to its ultimate goal, total readiness" (Army Family Action Plan III, DA Pam. 608-41, 1986, para 1-4c). There seems to be, indeed, general agreement that through helping soldiers solve some of their family problems, the Army is able to increase individual soldier and unit readiness as well as retention rates. However, to date there has been little hard evidence that such linkages do, in fact, exist. To determine the extent and direction of these linkages, the U.S. Army Research Institute for the Behavioral and Social Sciences (ARI) has initiated a comprehensive five-year investigation, the Army Family Research Program (AFRP).

A first step in establishing the connections between Army family programs and actions and soldier and unit readiness is to determine what constitutes readiness and how it can be measured. The Army, of course, currently assesses the readiness of its troops and units. The AFRP plans to take maximal advantage of such operational evaluations as Skill Qualification Tests and the Army Physical Readiness Test for individual soldiers, and Annual Command Readiness Inspections and Army Readiness Test and Evaluation Program results for units. In addition, we are trying to determine whether

---

[1]The views expressed in this paper are those of the authors and do not necessarily reflect the views of the U.S. Army Research Institute or the Department of the Army.

there are dimensions of readiness not covered by existing measures that should be incorporated into one or more composite indexes of readiness.

This report describes initial results of our investigation of soldier and unit readiness. Our overall approach is to measure readiness and relevant family factors through a variety of means (e.g., ratings, performance measures, and questionnaires) and through a variety of sources (e.g., enlisted personnel, officers, Army records, and spouses of soldiers). Composite readiness indexes, both at the individual soldier and unit (company-size) level, will then be formed on the basis of the inter-relationships among the component measures and the reliability and judged validity and relative importance of the measures.

Method

Workshops (Series 1)

Our first step in developing the readiness indexes was to conduct two series of four hour workshops. The first series of workshops was aimed at developing initial sets of individual soldier and unit readiness dimensions, and the second series was used to refine these dimensions. Each series consisted of eight workshops. The workshop participants consisted of 10 to 16 officers or NCOs from Combat Arms or Combat Service/Service Support units.

In the first series, the participants were given training in how to generate critical incidents that exemplified full readiness, lack of readiness, or some degree of readiness in between. The participants were instructed to record what actually happened that made them feel that the soldier or unit was or was not ready to perform the tasks required for successful accomplishment of the unit's mission. The participants first wrote critical incidents concerning the readiness of individual soldiers and then, after a break, incidents concerning the readiness of units (platoons, companies, or battalion level).

The critical incidents were analyzed in two phases. The critical incident data from the first four workshops, were numbered and independently reviewed by each of three researchers. Based on this review, each researcher generated a set of categories which he/she believed best represented a mutually exclusive and exhaustive list of dimensions of readiness. Once the dimensions were generated, the three researchers met to present their categories and discuss their rationales. Discussion proceeded until one set of mutually agreed upon dimensions was developed.

Following the generation of a common categorization scheme, each researcher independently categorized each critical incident from the workshops. Then the three researchers again met as a group to review those categories which were associated with low inter-rater agreement. The group discussed reasons for disagreement, confusion, or lack of clarity and the categories were revised to address these problems. Revisions included combining categories which appeared too closely related and adding further specification to categories which lacked clarity. The incidents were then reclassified into the final set of dimensions. Tables 1 and 2 list the dimensions obtained for soldier and unit readiness, respectively.

Table 1

Frequencies of Critical Incidents and Evaluations
of Preliminary Individual Soldier Readiness Dimensions

Soldier Readiness Dimensions	No. of Critical Incidents	No. of Times Selected Officers/NCOs	EP	No. of Comments on Rating Difficulty	No. of Comments on Rating Non-Applicability
1. Cooperation/Teamwork/Esprit' de Corps	5	91	102	3	2
2. Effort and Initiative	26	71	97	5	0
3. General Soldiering Skills	5	79	104	10	5
4. Improvement of Job Expertise	3	35	68	11	3
5. Individual Deployability (Army Task/Mission)	22	66	87	8	2
6. Individual Deployability (Personal/Family)	23	57	76	24	7
7. Job Discipline and Integrity	37	69	94	8	1
8. Job Technical Knowledge/Skill	29	96	95	6	1
9. Performance Under Pressure and Adverse Conditions	2	82	86	14	5
10. Personal Discipline	13	65	91	12	1
11. Physical Fitness and Health Maintenance	4	92	100	9	3
12. Relationship with Civilians in Host Country	1	6	8	42	21
13. Safety	31	35	53	4	3
14. Vigilance, Physical Security, and Handling Classified Materials	2	45	45	16	6
15. Care and Concern for Subordinates	5	97		6	3
16. Care and Concern for Subordinates' Families	5	57		10	6
17. Knowledge of Battlefield Tactics	2	67		23	22
18. Leadership	24	106		6	3
19. Maintaining Training Status of Subordinates	13	47		1	3
20. Relationships with Other Units	2	18		10	10
21. Assuring Unit Deployability	16	63		8	9
Total	270	1,344	1,106	236	116

Table 2

Frequencies of Critical Incidents and Evaluations
of Preliminary Unit Readiness Dimensions

Unit Readiness Dimensions	No. of Critical Incidents	No. of Times Selected	No. of Comments on Rating Difficulty	No. of Comments on Rating Non-Applicability
1. Adherence to Standards	15	87	8	3
2. Ammunition, Supplies, Materials, and Other Equipment (Not Including Vehicles and Weapons)	8	86	6	9
3. Care and Concern for Families	3	59	14	6
4. Care and Concern for Soldiers	9	93	6	2
5. Cohesion and Teamwork	30	95	4	1
6. Communication Within Unit	7	86	4	0
7. Cooperation/Coordination with Other Units	2	43	8	7
8. Emergent Leadership	10	55	14	5
9. Higher Echelon Support (Brigade, Battalion Level)	7	50	11	10
10. Leadership	36	109	6	1
11. Mission Performance	32	98	2	1
12. Personnel Capabilities	14	64	7	3
13. Personnel Deployability	8	51	0	4
14. Physical Fitness Program	6	85	3	6
15. Physical Security/Vigilance	3	30	5	2
16. Training Program	20	91	7	2
17. Unit Weapons	3	94	10	12
18. Vehicles/Transportation	8	83	29	18
Total	221	1,359	144	92

468

The critical incidents generated in the remaining four Series 1 workshops were used as a check to make sure that no new dimensions were needed and to help in the generation of behavior summary scales describing different degrees of readiness along the obtained dimensions. To construct the scales, behavioral descriptions for each dimension were written at three readiness levels (low, medium, and high) following a procedure used by Borman and Rose (1986).

Workshop Series 2

There were 115 participants altogether in the second series of eight workshops. They were first given a brief training session on the types of errors (e.g., halo) that raters often make. Then they were asked to rate three soldiers whom they felt most qualified to rate. The ratings for the 21 soldier readiness dimensions given in Table 1 were made on a seven-point scale. After completing the ratings, the participants were asked to select the 12 dimensions for officers and NCOs and the 10 dimensions for junior enlisted personnel (EP) that together would make the most comprehensive overall measure of readiness. The participants were also asked to indicate whether they had any difficulty rating individuals on the dimensions and whether there were certain types of soldiers for whom any of the readiness dimensions were non-applicable.

After a break, the participants were asked to repeat the rating and scale evaluation process on the three units (platoon and/or companies) that they felt most qualified to rate, using the unit readiness dimensions given in Table 2. The participants were asked to select the 12 out of 18 unit readiness scales that they thought would yield the most comprehensive measure of unit readiness. They were also asked to indicate whether they had any difficulty rating the units on the dimensions and whether there were certain types of units for which any of the readiness dimensions were non-applicable.

Results

Tables 1 and 2 present for the soldier and unit readiness dimensions, respectively, (a) the number of incidents obtained for each dimension in the first four workshops; (b) the number of times each dimension was selected to be included in the overall measure of readiness by the second series of workshop participants; (c) the number of comments the participants made concerning the difficulty of rating soldiers or units on the dimensions; and (d) the number of comments they made concerning the non-applicability of the scales to particular types of soldiers or units.

The evaluations made by the officers and NCOs clearly point to the possible elimination or modification of some scales and the retention of others. For example, individual readiness dimensions for Relationships with Civilians in Host Country and Relationships with Other Units were selected relatively few times for inclusion in an overall readiness index and were cited more frequently as being non-applicable to certain types of soldiers. On the other hand, judging from the number of times they were selected for inclusion in the overall index and the low number of non-applicability comments received, Physical Fitness and Health Maintenance and Cooperation/Teamwork/ Esprit 'de Corps are more universally relevant readiness dimensions. At the unit level, Physical Security/Vigilance was

considered a relatively unimportant readiness dimension. Leadership and Mission Performance, on the other hand, were often selected as relevant and seldom considered difficult to rate or non-applicable.

Apparently, the number of critical incidents obtained for each of the readiness dimensions may constitute a measure of the relative importance of the dimension in forming an overall readiness index. The number of critical incidents obtained for each of the dimensions in the first series of workshops was significantly negatively correlated with the number of comments on rating non-applicability obtained in the second workshop series ($r = -.47$ and $-.49$ for the soldier and unit readiness dimensions, respectively). Number of comments concerning rating difficulty for the dimensions were also negatively related to the number of critical incidents obtained, although the correlations ($-.36$ and $-.27$) were not significant. The number of times the officers and NCOs selected the dimensions for inclusion in an overall readiness measure was positively related to the number of critical incidents received for the dimension, especially in the case of units ($r = .62$). These correlations provide support for the use of the critical incident technique to identify readiness dimensions.

Discussion

Further analyses of the workshop data are planned prior to making final decisions concerning the readiness dimensions. Of particular interest will be the amount of overlap among the dimensions which can be assessed through correlational and factor analyses of the rating data. We do, of course, plan to use considerably fewer scales than the ones initially constructed while still maintaining much of their overall comprehensiveness. We had also hoped to have sets of soldier and unit readiness scales that could be applied to all Army company or platoon-sized units and their personnel. The comments received by the workshop participants, however, suggest the need for different sets of scales for different types of units and personnel. Having well-maintained vehicles is, for example, much more relevant to readiness for certain types of units than for others. Similarly, knowledge of battlefield tactics is much more important for some types of officers and NCOs than others.

We currently plan additional analyses and workshops to further refine and specify the dimensions of Army soldier and unit readiness. The rating procedures and scales will then be field tested prior to large scale data collections through which the linkages between readiness and family factors and Army community services will be determined.

REFERENCE

Borman, W.C., & Rose, S.R. (1986). Development of the Army-wide rating scales and task dimensions (ARI Technical Report 716). In Pulakos, E.D., & Borman, W.C. (Eds.), Development and field test of army-wide rating scales and the rater orientation and training program (pp. 8-16). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Basic Interests as Predictors of Attrition

Michael D. Matthews
Drury College
Springfield, Missouri 65802

Abstract

The Vocational Interest-Career Examination (VOICE) is an instrument designed to assess interests for vocational activities typical of work tasks found in Air Force enlisted occupations. Item responses on the VOICE are converted into 18 Basic Interest Scale (BIS) scores, each of which reflects interest in a relatively homogeneous content area. The present study examined the relationship among VOICE BIS scores, background variables, and Air Force first-term attrition in a sample of 44,487 first-term airmen. Results indicated that the BIS scores added significantly to the prediction of attrition. Theoretical and practical implications of the results are discussed.

Alley and Matthews (1982) described the development and psychometric characteristics of the Vocational Interest-Career Examination (VOICE), an instrument designed to assess interests for various vocational activities typical of work tasks found in Air Force enlisted occupations. Basic Interest Scale (BIS) scores derived from the VOICE have been found to be predictive of reported job satisfaction among Air Force first-term enlistees (Alley, Wilbourn, & Berberich, 1976). From knowledge concerning the relationships between BIS scores and subsequent job satisfaction, predicted job satisfaction (PJS) scores were developed for 20 groupings of related Air Force enlisted job specialties. Because job satisfaction is an important variable in models of employee turnover (e.g., Mobley, Griffeth, Hand, & Meglino, 1979) and individual work performance (.e.g., Seashore & Taber, 1975), VOICE PJS scores have been used to predict various occupational behaviors, including first-term attrition (Matthews, 1983), job performance (Berry & Matthews, 1983), and Air Force technical training attrition (Matthews & Ballentine, 1983). In general, VOICE PJS scores have been found to be significant predictors of these occupational outcomes, and to add significantly to the predictive power of background variables.

The studies noted above utilized VOICE BIS scores only indirectly, via VOICE PJS scores, to predict occupational behaviors. Although there are practical and conceptual grounds for using PJS scores to predict different criteria (Alley & Matthews, 1983), it may be that using VOICE BIS scores directly to predict occupational behaviors may result in more robust predictive relationships than when PJS scores are used as predictors. The purpose of the current research was to examine the relationship among VOICE BIS scores, background variables, and Air Force first-term attrition. The research examined the same sample used by Matthews (1983) and, with the exception of the substitution of BIS scores for PJS scores, similar regression models were used to predict attrition. Therefore, an indirect comparison of the relative efficacy of BIS and PJS scores in predicting attrition can be made by comparing the multiple Rs obtained for corresponding models in the two studies.

Method

Subjects

The sample consisted of 32,844 male and 11,643 female 1973-1975 enlistees. The subjects were typical of past Air Force accessions. Their average age was 18, the racial composition of the sample was similar to that of the United States population as a whole, and most (95 percent) had completed high school.

The VOICE

The VOICE consists of a 300-item vocational interest inventory requiring approximately 30 minutes to administer. Individual items are presented in booklet form and consist of occupational titles, work tasks, leisure time activities, and desired learning experiences. Respondents indicate relative preferences for each item in a standard like-indifferent-dislike (LID) format. Item responses were converted to two types of scales: (a) basic interest scales, and (b) occupational scales. The basic scales represent measures of general interest in various occupational and technical areas. They were constructed by grouping items of similar content into 18 independent sets covering a wide range of interests in the vocational and technical domain. The basic interest scales cover areas of Office Administration, Electronics, Heavy Construction, Science, Outdoors, Medical Service, Aesthetics, Mechanics, Food Service, Law Enforcement, Audiographics, Mathematics, Agriculture, Teacher/ Counseling, Marksman, Craftsman, Drafting, and Automated Data Processing. All items within each scale are homogeneous in the sense that each was selected to measure the same underlying dimension. The Office Administration items, for example, measure interest in clerical, administrative, and business related activities.

The occupational scales were designed for use in evaluating job assignment alternatives. It has been found that certain patterns of basic interest scores predict job satisfaction in various Air Force job clusters (Alley, et al, 1976). These clusters, 20 in number, represent an exhaustive categorization of Air Force job specialties. The VOICE occupational scales, therefore, provide a PJS score for each of these 20 job clusters. Consequently, if used operationally, job placement personnel would be able to readily obtain a prediction of job satisfaction for any Air Force career field by determining in which of the clusters that particular job falls. The occupational scales, while formulated from basic interests, provide direct estimates of job satisfaction for each career field in the set and can be used for making specific comparisons between alternative assignments (Alley et al, 1976). PJS scores range from 200 to 800, with a mean of 500 and standard deviation of 100. For a more thorough and technical discussion of the development of the VOICE and a description of the basic interest and occupational scales, their psychometric characteristics, and validity, see Alley and Matthews (1982).

Procedure

The VOICE was administered to the sample during Basic Military Training. The subjects were allowed to complete their initial four to six year duty obligation, and attrition rates were determined. Attrition was defined as failing to complete at least 36 months of the initial active duty obligation.

Information on Armed Services Vocational Aptitude Battery (ASVAB) scores, Armed Force Qualification Test (AFQT) scores, age, and education level were also obtained for each subject.

Results and Discussion

Table 1 presents the results of an overall regression analysis designed to assess the relationship among VOICE BIS scores, background variables, and first-term attrition. Besides the 18 VOICE BIS scores, variables entered into the regression included background variables (i.e., ASVAB aptitude index scores, AFQT scores, age, and education level), and group membership (i.e., which of the 20 VOICE occupational groups into which a recruit was assigned). Comparison 1 indicates that knowledge of group membership is a significant predictor of attrition. Comparison 2 shows that BIS scores and background variables together add significantly to the predictive power of group membership. Similarly, background variables alone (Comparison 3) and BIS scores alone (Comparison 4) are predictive of attrition over and above the effects of group membership. Finally, and importantly, Comparison 5 shows that BIS scores add significantly to the prediction of attrition, above and beyond the influence of background variables and group membership.

A similar pattern of results was found when VOICE PJS scores were used to predict attrition. Matthews (1983) reported that PJS scores alone were significant predictors of attrition, and added significantly to the predictive power of background variables and knowledge of group membership. However, the multiple Rs reported by Matthews (1983), although significant, were smaller than those found for corresponding models in the current study. For the full model (PJS Scores + background variables X group membership), Matthews obtained a multiple R of .142, compared to a multiple R of .234 for the corresponding model (i.e., BIS + Background X group membership) in the current study. For the VOICE PJS score alone model, Matthews obtained a multiple R of .080, compared to a value of .193 for the BIS alone model in the current study. Thus, it would appear that attrition may be predicted somewhat more effectively when BIS scores are used directly in estimating attrition, rather than when they are used indirectly via PJS scores.

The results emerging from the VOICE research program have important implications for the Air Force's personnel classification system, and are also of conceptual interest. Each year, the Air Force recruits approximately 50,000 enlisted persons, of whom in excess 30 percent (i.e., 15,000 or more) may be expected to fail to serve out at least three years of their initial duty obligation. Each attrition is, of course, quite costly to the Air Force in terms of both dollars and loss of expertise. Because of the large numbers of personnel involved, even a small percentage reduction in attrition would translate into substantial savings for the Air Force. At a conceptual level, the results of the VOICE research program underscore the complexity of the relationships among interests, job satisfaction, and turnover, and suggest that vocational interests may play a more central role in influencing turnover that is usually attributed to them.

Future work in this area should use BIS scores to predict other criteria, including technical school attrition, job performance, and retention.

Preliminary data from the VOICE research program indicates that BIS scores may provide a better prediction of retention than do PJS scores. Further research may also examine more closely the effects of gender occupational-group interactions on attrition and other occupational behaviors. Additionally, the results of Matthews (1983), Berry and Matthews (1983), and those of the current study could be integrated to produce a personnel classification algorithm that would optimally mix interests, predicted job satisfaction, aptitudes, and other background variables to result in a more refined placement of individuals into jobs. Finally, much of this work could be extended to examine the relatinships among interests, job satisfaction, and occupational behaviors in non-military occupations.

REFERENCES

Alley, W. E., & Matthews, M. D. The Vocational Interest-Career Examination: A description of the instrument and possible applications. Journal of Psychology, 1982, 112, 169-193.

Alley, W. E., & Matthews, M. D. (1983). Predicted Job Satisfaction and Attrition Due to Poor Work Performance. Paper presented at the 91st Annual Convention of the American Psychological Association, Anaheim, California.

Alley, W. E., Wilbourn, J. M., & Berberich, G. L. Relationships between performance on the Vocational Interest-Career Examination and reported job satisfaction. AFHRL-TR-76-89, AD-A040 754. Lackland AFB, TX: Personnel Research Division, Air Force Human Resources Laboratory, December 1976.

Berry, G. A., & Matthews, M. D. Predicting job satisfaction and job performance. Proceedings, 24th Annual Conference of the Military Testing Association. AFHRL-TP-83-16, AD-A126 554. Brooks AFB, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory, March 1983, pp 20-24.

Matthews, M. D. Vocational interests and job satisfaction: Effects on turnover among Air Force enlistees. Proceedings, 24th Annual Conference of the Military Testing Association. AFHRL-TP-83-16, AD-A126 554. Brooks AFB, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory, March 1983, pp 130-134.

Matthews, M. D., & Ballentine, R. D. (1983). Technical School Attrition as a Function of Predicted Job Satisfaction. Proceedings of the 25th Annual Military Testing Association Meeting. Gulf Shores, Alabama, Naval Personnel Research and Development Center.

Mobley, W. H., Griffeth, R. W., Hand, H. H., & Meglino, B.M. Review and conceptual analysis of the employee turnover process. Psychological Bulletin, 1979, 86, 493-533.

Seashore, S. E., & Taber, T. D. Job satisfaction indicators and their correlates. American Behavioral Scientist, 1975, 18, 333-368.

Table 1

Results of Overall Regression Analyses

	REGRESSION MODEL		MULTIPLE R				
Comparison	Full	Restricted	Full	Restricted	df1	df2	F*
(1)	Group	Unit Vector	.0400	.0000	20	44,466	44.45
(2)	(BIS + Background) X Group	Group	.2324	.1400	619	43,487	2.58
(3)	Background X Group	Group	.1954	.1400	259	44,207	3.30
(4)	BIS + Group	Group	.1926	.1400	359	44,107	2.23
(5)	(BIS + Background) X Group	Background + Group	.2324	.1954	360	43,487	2.06

X With degrees of freedom of these magnitudes, any F ratio above 1.000 is significant at the .05 level.

475

Pathological Gambling:
The Addiction of the Eighties

Valerie C. Lorenz, Ph.D.
National Center for Pathological Gambling, Inc.
Baltimore

Abstract

Pathological gambling is an impulse control disorder
leading to addictive behavior which ultimately results
in financial ruin, broken families, medical problems
and suicide attempts. Types of gamblers, prevalence of
compulsive gambling, and theories of this psychiatric
disorder are presented. Pathological gamblers are
generally highly motivated for treatment and respond
well to cognitive/rational-emotive therapy.

Gambling is prevalent on all military bases, whether it is
in the form of card playing, shooting dice, sports betting or
playing slot machines. Off base military personnel and their
dependents can go to race tracks, casinos, or card rooms, or they
can spend their money on lottery tickets. Compulsive gamblers
have a history of big wins, and this sudden pot of money and
attention makes them feel good, for a change. Losses become
emotionally intolerable, a loss of self-esteem and not just a
loss of money (Galski, 1987).

> Case 1: SrA. "Lance" had been in the Air Force (NSA)
> for one year, was given duty assignments beyond his
> rank and length of service because he applied himself,
> learned his duties quickly, and performed outstanding-
> ly. He learned to gamble on his first TDY when he was
> sent to Nellis AFB. His sergeant taught him how to play
> blackjack and within a few hours he won $2,000. Three
> days later, when his TDY was completed, he had written
> $300 worth of checks. He was determined to win it
> back. Two years later, at age 22, suffering from high
> blood pressure, severe depression, and generalized
> anxiety, he was facing a military court martial for
> falsifying travel vouchers to obtain money to feed his
> addiction (USAF v. SrA. XX, 1985 - Anonymity protected).

With the movement of the past decade to legalize various
forms of gambling throughout the United States and on overseas
military bases, gambling has not only become socially accept-
able, accessible, and available to almost anyone, but gambling
is also actively promoted as an act of civic responsibility.
"It is good for morale. Proceeds are used for family services"
is the rationale for slot machines on overseas Army bases
(Farinella, 1987).

Current trends indicate that the incidence of compulsive
gambling is steadily increasing. In 1976 the U.S. Commission
on the Review of the National Policy Toward Gambling determined
that .77% of the adult population in the United States, approxi-
mately 1.1 million people, could be identified as compulsive
gamblers. The incidence in Nevada was nearly double that figure
(U.S. Commission, 1976). More recent studies sponsored by state
health departments or lotteries estimated the incidence at over
3%. This does not include teenagers, who are also appearing in
greater numbers at Gamblers Anonymous and in professional treat-
ment settings.

Pathological gambling now has become a truly democratic
illness, in that it strikes young and old, male or female, people
of any ethnic background or religion, and from all socio-economic
groups. They may resort to only one form of gambling or to
several, may lose control within a month or may gamble respon-
sibly for several years before losing control. But the devasta-
tion is there in every case.

What, then, is pathological gambling? And how does patholo-
gical gambling differ from other types of gambling? What are the
theories of etiology of this psychiatric disorder?

What is Pathological Gambling?

Pathological gambling has been classified by the American
Psychiatric Association in its Diagnostic and Statistical Manual
of Mental Disorders, Third Edition and Third Edition, revised,
under Impulsive Control Disorder (Section 312.31). The person
is chronically and progressively unable to resist impulses to
gamble, to the point that gambling compromises, disrupts, or
damages family, personal, and vocation pursuits (1987).

Types of Gamblers

Basically, there are four types of gamblers. Most frequent
is the Social Gambler, who gambles for recreation or diversion
from everyday stresses. Losses are considered the cost of enter-
tainment, and gambling does not interfere with normal family,
social or vocational interests. Should gambling interfere, the
social gambler will limit the gambling or turn to an activity
which is less disturbing.

The Professional Gambler views gambling as a business.
The gambling is disciplined and controlled, with losses being
carefully studied to minimize their recurrence. The professional
gambler earns his livelihood from gambling.

The Criminal Gambler gambles to make money, even if this
includes cheating or swindling, alone or in conspiracy with
others. Losses are usually blamed on others and cheating is
justified. The criminal gambler generally has a history of anti-
social behavior since early childhood on.

The Pathological Gambler can be described as an individual who is above average in intelligence, competitive, energetic, hardworking, motivated to achieve, honest and law-abiding. The compulsive gambler has a solid set of values concerning law and order, family, health, job and country.

Typically the pathological gambler is raised in a family environment of strict but inconsistent discipline, with a strong emphasis on money or materialistic possessions, emotional or parental deprivation, and a history of physical, verbal, and/or sexual abuse. There is a history of compulsive gambling, alcoholism or some other serious psychiatric disorder in the family.

The compulsive gambler is emotionally immature, a "loner" with low frustration tolerance, who is easily bored, and whose self-image and self-esteem is at the minus-zero level. The compulsive gambler also has a history of unresolved traumas, experiences which trouble the gambler for many years. They are dysphoric people constantly seeking acceptance and approval from others (Taber & Boston, 1987).

Gambling gives them a sense of action and excitement, the escape from the pain in their lives. Winning gives them a sense of confidence and accomplishment, attention and acceptance from others. Big wins encourage irrational thinking that somehow they have superior gambling skills and luck. As the delusional thinking becomes more fully developed with increased gambling, the thought processes become marked by denial, self-deceptions, and obsessive thoughts of gambling (Gaboury & Ladouceur, 1987).

Theories of Compulsive Gambling Etiology

Not any one theory developed so far explains all compulsive gambling. A sociological theory developed by Lesieur (1977) explains how the gambler first uses up all legal options of funding, such as earnings, savings, and loans, and then resorts to illegal activities to acquire funds to pay off gambling debts and to continue gambling. Illegal activities most often are in a variety of bad checks, insurance fraud, forgery, and embezzlement. The gambler "hits bottom" when there is no more access to funds or when the gambler is about to be arrested.

Jacobs (1986) developed a theory of the Addictive Personality Syndrome which applies to compulsive gamblers. "Addiction is seen as a dependent state acquired over time by a predisposed person in an attempt to correct a chronic pre-existing stress condition" (p. 20). "The presence of an atypical and persistent physiologic state is held to be one of two necessary predisposing conditions for developing an addiction. The second essential pre-condition...is a childhood and adolescence marked by deep feelings of inadequacy, inferiority, and a sense of rejection by parents and significant others (page 21).

Taber & Fuller espouse a similar theory, that of a dysphoric childhood and adolescence which contributes to the vulnerability of individuals prone to addictive behaviors (1987). They hold it is irrelevant whether a substance is ingested or whether the addiction is manifested in behavior, such as compulsive gambling of sexual addiction. Taber also found evidence of thyroid disease among some of his compulsive gambling patients.

The possibility of biochemichal imbalance among compulsive gamblers was explored by researchers at the National Institutes of Health. Roy (1987) found that serotonin levels differed significantly between normals and compulsive gamblers.

In short, there is evidence to support theories that compulsive gambling may be the result of genetic transmissior, due to biochemical imbalance, a function of learning, a means of seeking pleasure while avoiding pain, and a consequence of having the right personality characteristics.

Contrary to popular belief, compulsive gamblers do not gamble to lose - they need to win in order to enhance their self-esteem and to continue in the action. Nor do they gamble for greed. They want love and acceptance from others, even though at the same time their experiences have taught them to shy away from intimacy and to be distrustful.

> Case 2. Richard P. graduated from law school,
> expanded the legal offices at the Army base from
> one office to a twelve-office building, and was
> sent to Vietnam. His classmate was killed the
> first week he was in that country, his helicopter
> was shot down three times, and on two occasions
> he signed orders only to have the men killed in
> action. He started shopping compulsively, and
> upon return to his law practice started going
> to the casinos and race tracks. His family did
> not learn of his Vietnam experiences until he
> was facing criminal charges for having violated
> his clients' escrow accounts.

> Case 3. Roy Q. served in the U.S. Navy for eight
> years, having joined up after his marriage ended
> in divorce. His mother, a status conscious member
> of the local country club, was embarrassed that
> he was an enlisted man rather than an officer.
> Bob loved the navy, and planned to make that his
> career, until after a particularly gruesome assign-
> ment, when he had to "fish bloated arms and pieces
> of bodies out of the ocean." He never told anyone
> about his nightmares. Instead, he would try to
> block out his memories with all-night poker games.
> When he presented for treatment, he was suicidal.

Treatment

Compulsive gamblers rarely present with only one disorder, pathological gambling. Almost always they can also be diagnosed as suffering from major affective disorders, such as dysthymia or bipolar disorders. Generalized anxiety is another frequent diagnosis. In Case 2 and Case 3 obviously Post Traumatic Stress Syndrome was as much a focus on treatment as the gambling.

Treatment that works effectively with compulsive gamblers is cognitive and rational-emotive therapy, a la Beck and Ellis, chipping away at the thought distortions and irrational beliefs which make up the gambler's cognitions. In addition, gamblers require individual therapy. These are traumatized people who do not share their burdens with others; however, a skillful therapist who has established good rapport can break down the defenses and bring the pain and anger out in the open. Then the healing process can begin, with individual work, in groups, and GA.

References

American Psychiatric Association. (1987). Diagnostic and statistical manual of mental disorders. Third edition, revised. Washington, DC: APA.

Fuller, N.P. & Taber, J. (1987). On the irrelevance of substances in defining addictive disorders: Pathological gambling. Paper presented at the 7th International Conference on Gambling and Risk Taking, Reno.

Gaboury, A. & Ladouceur, R. (1987). Irrational thinking and gambling. Paper presented at the 7th International Conference on Gambling and Risk Taking, Reno.

Galski, T. (ed.) (1987). The handbook of pathological gambling. Springfield, IL: Charles C. Thomas.

Jacobs, D. F. (1986). A general theory of addictions: A new theoretical model. Journal of Gambling Behavior, 2(1), 15-31.

Lesieur, H. (1977). The chase. New York: Anchor.

Roy, A. (1987). A search for biological substrates to pathological gambling. Paper presented at the 7th International Conference on Gambling and Risk Taking, Reno.

Taber, J. & Boston, M.D. (1987). Developmental vulnerability in the etiology of problem gambling and other addictions. Paper presented at the 7th International Conference on Gambling and Risk Taking, Reno.

US Commission on the Review of the National Policy Toward Gambling. (1976). Final Report. Washington, DC: USGPO.

The Relationship Between Job Enrichment
and Organizational Commitment
Among Air Force Personnel

Robert E. Niebuhr
Carla R. Hodges
Auburn University

As originally conceived, the job characteristics model of
Hackman and Oldham (1976) included the individual characteris-
tics of the task and several outcome variables. Hackman and
Oldham assumed that one of the most important work values is
the job incumbent's need for personal growth and development
through his or her job. Employees with high growth need
strength (GNS) should respond more positively to jobs that
have high levels of the five core task dimensions (task sig-
nificance, skill variety, etc.) than employees with low GNS.
In a recent meta-analysis of 28 job enrichment studies, Loher,
Noe, Moeller, and Fitzgerald (1985) found a correlation
between task characteristics and job satisfaction of .68 for
persons who were high on GNS and .38 for persons who were low
on GNS.

The purpose of the study reported in this paper was to
examine the relationship between task characteristics and
organizational commitment, and the impact which several moder-
ator variables have upon this relationship. Based upon evi-
dence by Steers (1977) and Hunt, Chonko, and Wood (1985) there
appears to be a positive relationship between enriched tasks
and organizational commitment. Because of the source of data
used in this study, five demographic variables were chosen as
potential moderators of the job enrichment - organizational
commitment relationship: sex, race, marital status, age, and
educational level.

In seeking to better understand these relationships, the
following hypotheses were addressed:

1. There is a positive correlation between a perceived
   enriched job and organizational commitment.
2. The relationship between perceived job enrichment and
   organizational commitment will be moderated by sev-
   eral individual demographic variables (i.e., sex,
   race, marital status, age, educational level).

RESULTS

For this particular study, a 37,838 person segment was
chosen from the Air Force's Organizational Assessment Package
(OAP) data base of more than 200,000 responses to provide a
representative sample of various functional area categories
and personnel classifications (i.e., officer, enlisted, civil-
ian).

Table 1 shows that perceived job enrichment is positively correlated with age, educational level, and organizational commitment. The correlation between job enrichment and organizational commitment was significant. This finding supports hypothesis 1 and previous studies reviewed earlier which suggested that individuals with more enriched tasks would have greater organizational commitment.

TABLE 1

Descriptive Statistics and Intercorrelations Among

The Continuous Study Variables[a,b]

	M	SD	1	2	3
1. Perceived Job Enrichment	110.60	66.58			
2. Organizational Commitment	24.74	7.09	.411		
3. Educational Level	3.23	1.25	.140	.163	
4. Age	29.87	8.57	.287	.212	.238

[a]Sample sizes ranged from 34949 to 37681 for most of the variables.

[b]All correlations were significant ($p<.001$)

Analyses involving moderator variables were conducted by using moderated regression analysis. The moderator variables and perceived job enrichment were used in five separate moderated regression equations in order to predict organizational commitment. For categorical variables, (e.g., marital status) dummy variables were created to allow the regression analysis to be accomplished. Table 2 presents the results of these analyses.

The only moderator which did not significantly affect the job enrichment/organization commitment relationship was race. The moderated regression analysis did, however, provide enough support for hypothesis 2 which suggested that some of these individual variables should have a significant moderating impact on the relationship.

TABLE 2

Results of Moderated Regression Analysis For Moderators of the

Job Enrichment Organizational Commitment Relationship

Dependent Variable	Regression Variables	Interaction Effects		
		R^2	ΔR^2	F
ORGC	JE, AGE, JExAGE	.177	.00100	81.60*
ORGC	JE, RACE, JExRACE	.169	.00001	.41
ORGC	JE, SEX, JExSEX	.168	.00060	25.61*
ORGC	JE, EDL, JExEDL	.179	.00014	5.85*
ORGC	JE, MARS, JExEDL	.169	.00013	5.33*

Notes: df=1,34006; ORGC=Organizational Commitment, JE=Job Enrichment, EDL=Education Level, MARS=Marital Status

*R<.001

DISCUSSION

This study attempted to use enriched tasks as a predictor variable for organizational commitment as moderated by five individual variables. The results indicated that a positive relationship exists between enriched tasks and organizational commitment for this sample population. Moderated regression analysis also indicated that significant interactions occurred when examining several individual demographic variables. By dichotomizing the data and analyzing subgroups it was found that a stronger relationship between perceived job enrichment and organizational commitment does exist for older employees (r=.40) versus younger employees (r=.36) and for married individuals (r=.41) versus single employees (r=.38) and for males (r=.42) versus females (r=.34). Individuals who had attained a college degree were found to have a stronger job enrichment-enrichment-organizational commitment relationship (r=.43) than those individuals who did not (r=.40). Race apparently did not moderate the relationship between job enrichment and organizational commitment for these subjects.

Two limitations are suggested for the results found in this study. The correlation matrix for the study indicated multicollinearity (Kerlinger & Pedhazur, 1973) and the results also need to be interpreted with regard to the population studied. Too many of the variables were intercorrelated, a result often occurring in attitudinal studies using all self-reported data. The large sample size, however, does provide the advantage of being able to verify that many of relationships were significant despite these intercorrelational prob-

lems. With regard to the generalizability of these results,
it must be realized that the questionnaire was developed for
Air Force personnel and that few other institutions in society
conscientiously stress and attempt to perpetuate organiza-
tional values, goals, and attitudes to the degree the military
does (Sarkesian, 1972). The military encourages its personnel
to obtain more education whether it be through military
schools or through more formalized college courses supported
by tuition matching plans. Frequently, personnel are rewarded
upon successful completion of additional education with an
increase in pay and/or a promotion. This type of policy does
provide an explanation for the finding that education had a
positive rather than a negative influence on the job
enrichment-organizational commitment relationship. Certainly
the fairly autonomous nature of the military puts it in a gen-
erally stronger position to affect the beliefs of its members
(Lang, 1972).

This study as well as previous studies show that continued
research on the task characteristics-organizational commitment
relationship is needed. Enriching tasks appears to be a legi-
timate path in developing employee satisfaction and commitment
to the organization. The results of attempts to improve orga-
nizational commitment (and, possibly, other dependent vari-
ables) through job design are also likely to have greater
results with different segments of the work force, as evi-
denced by the results of this particular study.

REFERENCES

Hackman, J. R. and Oldham, G. R. 1976. Motivation
through the design of work: Test of a theory, _Organiza-
tional Behavior and Human Performance_, 16: 250-279.

Hunt, L. B. Chonko, & Wood, V. R. 1985. Organizational
commitment and marketing, _Journal of Marketing_, 49:
112-126.

Kerlinger, F. N. & Pedhazur, E. J. 1973. _Multiple
regression in behavioral research_. New York: Holt,
Rinehart, & Winston.

Lang, K. 1972. Military career Structure: Emerging
trends and alternatives, _Administrative Science Quar-
terly_, Vol. 17 (December 1972), 49.

Loher, Brian T., Noe, R. A., Moeller, N. L., Fitzgerald,
M. P. 1985. A meta-analysis of job characteristics
to job satisfaction, _Journal of Applied Psychology_, 70:
280-289.

Sarkesian, S. C. 1972. Political soldiers: Perspectives
on professionalism in the U.S. Military, _Midwest Jour-
nal of Political Science_, 16: 239.

Steers, R. M. 1977. Antecedents and outcomes of organi-
zational commitment. _Administrative Science Quarterly_,
22: 46-56.

Table 1 shows that perceived job enrichment is positively correlated with age, educational level, and organizational commitment. The correlation between job enrichment and organizational commitment was significant. This finding supports hypothesis 1 and previous studies reviewed earlier which suggested that individuals with more enriched tasks would have greater organizational commitment.

TABLE 1

Descriptive Statistics and Intercorrelations Among

The Continuous Study Variables[a,b]

		M	SD	1	2	3
1.	Perceived Job Enrichment	110.60	66.58			
2.	Organizational Commitment	24.74	7.09	.411		
3.	Educational Level	3.23	1.25	.140	.163	
4.	Age	29.87	8.57	.287	.212	.238

[a]Sample sizes ranged from 34949 to 37681 for most of the variables.

[b]All correlations were significant ($p < .001$)

Analyses involving moderator variables were conducted by using moderated regression analysis. The moderator variables and perceived job enrichment were used in five separate moderated regression equations in order to predict organizational commitment. For categorical variables, (e.g., marital status) dummy variables were created to allow the regression analysis to be accomplished. Table 2 presents the results of these analyses.

The only moderator which did not significantly affect the job enrichment/organization commitment relationship was race. The moderated regression analysis did, however, provide enough support for hypothesis 2 which suggested that some of these individual variables should have a significant moderating impact on the relationship.

TABLE 2

Results of Moderated Regression Analysis For Moderators of the

Job Enrichment Organizational Commitment Relationship

Dependent Variable	Regression Variables	Interaction Effects		
		R^2	ΔR^2	F
ORGC	JE, AGE, JExAGE	.177	.00100	81.60*
ORGC	JE, RACE, JExRACE	.169	.00001	.41
ORGC	JE, SEX, JExSEX	.168	.00060	25.61*
ORGC	JE, EDL, JExEDL	.179	.00014	5.85*
ORGC	JE, MARS, JExEDL	.169	.00013	5.33*

Notes: df=1,34006; ORGC=Organizational Commitment, JE=Job Enrichment, EDL=Education Level, MARS=Marital Status

*R<.001

DISCUSSION

This study attempted to use enriched tasks as a predictor variable for organizational commitment as moderated by five individual variables. The results indicated that a positive relationship exists between enriched tasks and organizational commitment for this sample population. Moderated regression analysis also indicated that significant interactions occurred when examining several individual demographic variables. By dichotomizing the data and analyzing subgroups it was found that a stronger relationship between perceived job enrichment and organizational commitment does exist for older employees (r=.40) versus younger employees (r=.36) and for married individuals (r=.41) versus single employees (r=.38) and for males (r=.42) versus females (r=.34). Individuals who had attained a college degree were found to have a stronger job enrichment-enrichment-organizational commitment relationship (r=.43) than those individuals who did not (r=.40). Race apparently did not moderate the relationship between job enrichment and organizational commitment for these subjects.

Two limitations are suggested for the results found in this study. The correlation matrix for the study indicated multicollinearity (Kerlinger & Pedhazur, 1973) and the results also need to be interpreted with regard to the population studied. Too many of the variables were intercorrelated, a result often occurring in attitudinal studies using all self-reported data. The large sample size, however, does provide the advantage of being able to verify that many of relationships were significant despite these intercorrelational prob-

Construction and Preliminary Validation
of an Equal Opportunity Climate Assessment Instrument

Dan Landis and Gloria Fisher
University of Mississippi

Mickey R. Dansby
Defense Equal Opportunity Management Institute

Abstract

Construction and preliminary validation of an instrument to assess equal opportunity climate in the military were begun. The research was conducted at the Defense Equal Opportunity Management Institute (DEOMI), Patrick Air Force Base, FL. Students undergoing equal opportunity training served as subjects. A definition of Equal Opportunity Climate and a model linking Equal Opportunity Climate to other organizational variables are presented. Preliminary results in the development of the instrument indicate it is reliable and has some measure of construct validity. Further laboratory research and field validation at a random selection of military bases are recommended.

Although previous research on climate in military organizations has not focused on the construct of Equal Opportunity (EO) Climate, several researchers have attempted to assess both organizational climate and race relations climate in the military. Bowers (1975) measured organizational climate variables in the Navy by using the Survey of Organizations (SOO). In general he found that 'on all measures of organizational climate . . . Navy respondents were lower than nearly three-fourths of civilian respondents.' The findings revealed more felt discrimination among minorities—particularly Blacks—and at the same time showed a negative relationship between felt discrimination and climate (i.e., the better the climate, the less the felt discrimination). In another Navy study, Parker (1974) found almost no difference between races in perceptions of organizational climate. Results indicated racial composition of the work group was a critical moderator variable of the relationship between experienced practices and felt discrimination. Research by Pecorella (1975) indicated anizational climate measures presented patterns of (if anything) perceived reverse discrimination (although ective data, such as advancement and training opportunities did not). Pecorella also noted that felt rsonal discrimination seems to be closely tied to one's immediate work environment (particularly to advancement opportunities and friendly relations with one's peers). This research suggests much of the perception that one is discriminated against stems from job characteristics (e.g., promotions) and relations with one's co-workers.

In surveys by the Army Research Institute (Brown, Nordlie, & Thomas, 1977) there was a notable difference in how the 'race problem' was seen by Whites and Blacks in the Army. Whites in the Army tended to accept the proposition that the Army is free from racial discrimination. Blacks saw the Army as highly discriminatory. This difference also correlated with grade. Officers and higher enlisted saw the race problem as less serious than did the lower enlisted grades. The 1972 results were replicated in 1974, in spite of the existence of an all-volunteer Army and an increase in Black enlisted individuals. In 1978, Hiett and Nordlie concluded (in their study on unit race relations program in the Army) that despite the relative absence of overt interracial violence, race-related tensions persist.

Most research on climate and race relations in the military has focused on differences between Blacks and Whites. That focus has now been expanded to include other racial/ethnic minorities, religious discrimination and sexual discrimination and harassment. In one of the few research efforts in the military regarding sexual harassment, a survey of 104 Navy women (Reily, 1980), almost all had experienced sexual harassment in their careers; lower grade enlisted women were harassed the most. The data indicated sexual harassment negatively affected the attitude of the female service member, as well as her desire and intent to remain.

The military could benefit by a reliable and valid instrument which could be used, along with other data such as objective management indices, to effectively assess Equal Opportunity Climate. The present research was conducted to provide a definition of Equal Opportunity Climate, begin construction and preliminary validation of an instrument to measure Equal Opportunity Climate in all Services, and hypothesize a model which relates equal opportunity climate to other organizational variables. The instrument under development has been tentatively named the Military Services Equal Opportunity Climate (MISEOC) survey.

<u>Method</u>

Definition

Equal Opportunity Climate is defined as

the expectation by individuals that opportunities, responsibilities, and rewards will be accorded on the basis of a person's abilities, efforts, and contributions; and not on race, color, sex, religion, or national origin. It is to be emphasized that this definition involves the individual's perceptions and may or may not be based on the actual witnessing of behavior.

Model of Equal Opportunity Climate

The model (Figure 1) developed in this project is an expansion of the definition given above. EO Climate is seen as the result of several cognitive operations, which in turn have antecedents both in the person's past history and in events in the outside world. At the same time, EO Climate has impacts on a number of cognitive and motivational processes which are part of what may be called 'readiness.' Thus, EO Climate is essentially an internal process which is only tangentially under the control of environmental events. As such, the model suggests that command's responses must be timely and vigorous in order to change negative or reinforce positive expectations by personnel.
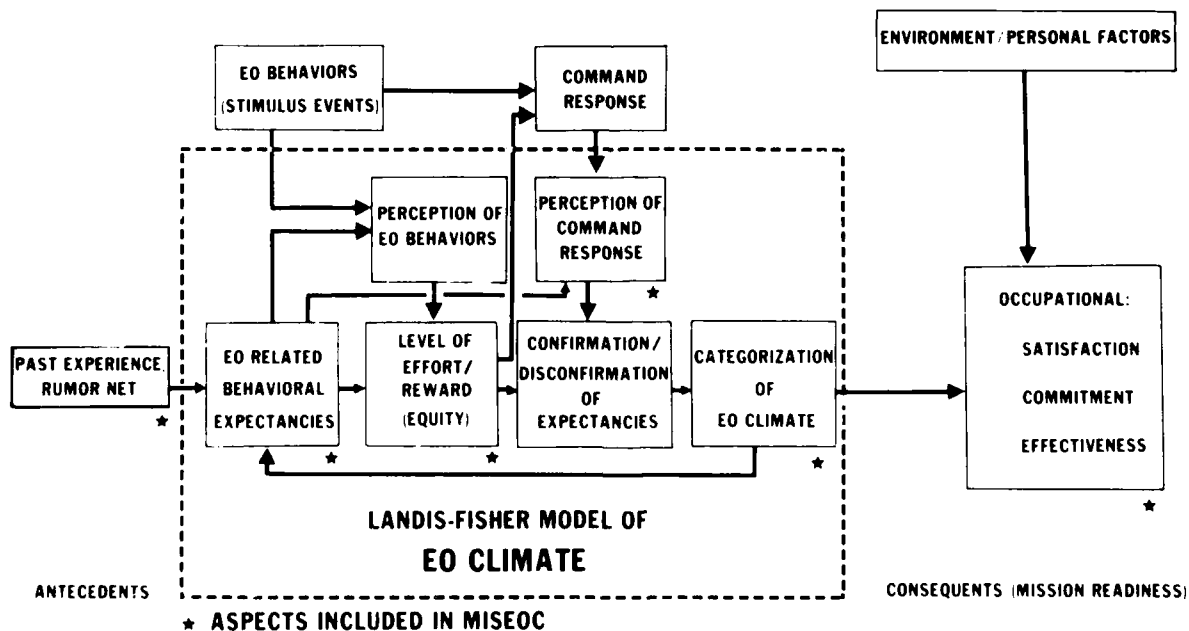
Figure 1. Model of Equal Opportunity Climate and Readiness.

Sample

The respondents were the members of DEOMI class 87-2 (51 Men, 11 Women; 6 Officers, 57 Enlisted Personnel; 42 Army, 5 Air Force, 14 Navy, 2 Undefined. The total is less than the number used for analysis because 11 respondents provided no identifying information).

Questionnaire Design

Elicitation and Ranking of Behaviors. The academic/training staff of DEOMI (n=20) were asked to list 'five specific behaviors that would be indicative of 'poor equal opportunity climate.'' An example of a behavior was provided. Emphasis was on a definable unit of behavior, not a series of events spread out over time. Over 100 behaviors were suggested. Analysis for redundancy and specificity reduced the list to 78. A separate group of Guard and Reserve personnel (n=50), taking a DEOMI short course, was asked to assign a rating of 1-10 to each behavior. The low end of the scale represented 'No importance,' while the high end was anchored at 'Of critical importance.' Further examination and analysis reduced the number of items to 71.

Design of Response Dimensions. Responses were solicited on 5-point scales. For the Equal Opportunity Behaviors section, the scale anchors were

1=There is almost no chance that the behavior occurred.
2=There is a small chance that the behavior occurred.
3=There is a moderate chance that the behavior occurred.
4=There is a reasonably high chance that the behavior occurred.
5=There is a very high chance that the behavior occurred.

Items for Measuring Work-group Effectiveness and Organizational Commitment. Though not analyzed in the present research, items for these two scales were included for future research. For Work-group Effectiveness, 15 items from the United States Air Force Organizational Assessment Package (Short, 1985) were selected. These items had high loadings on factors labeled Work Group Effectiveness and General Organizational Climate. For Organizational Commitment, 15 items from Mowday, Steers, and Porter (1979) were selected and rewritten to conform to a military situation.

Development of Situational Scenarios. In order to test the MISEOC format for discriminant validity, two hypothetical locales were devised. Information was provided on each locale along six dimensions taken from the management indices used by the United States Air Force to assess the level of human relations climate (Air Force Pamphlet 30-13, 21 January 1985). The six dimensions were percentages of Articles 15, involuntary separations, and courts-martial for minorities and Whites; discrimination and sexual harassment complaints filed and confirmed; and hate group activities in the local area. The locales were described on each dimension as having 'an above average rate' or 'a significant change from the previous year' for the 'poor' EO Climate locale. For the 'good' locale, the descriptors were 'a below average rate' or 'a significant reduction from the previous year' on each dimension.

Experimental Design

Variations of the questionnaires were administered twice to class 87-2.

Administration 1. Lists of the 71 Equal Opportunity Climate behaviors were given to 74 members of the class. The students were to judge the 'chances that the behavior occurred in the non-DEOMI portion of Patrick Air Force Base during the past 30 days.' This administration was used to explore the factorial structure of each part of MISEOC and to assess the reliability of each factor and the total survey.

Administration 2. This was seven days after the first survey. Again, the target of Patrick Air Force Base was used; however, the order of the behaviors was randomly mixed from the first administration. In addition, half of each group was asked to rate the good locale and half the poor locale on each of the 71 behaviors. A manipulation check was added by asking at the end of group of the 71 behaviors how 'most people at this locale

would rate the equal opportunity climate.' The response was made on a 1-5 scale from 'very poor' to 'very good.' The second administration was used to assess test-retest reliability and probe the discriminant validity of MISEOC.

Results

Separate Principal Components Factor Analyses of the 71 equal opportunity behaviors were performed for each administration of the Patrick and the good and poor versions. The analyses used unity in the diagonals and Varimax rotation (oblique rotation tailed to converge in 25 iterations). Cronbach's alphas were computed for the total and for factors of each administration. In addition, scores were computed for the total and for factors from each administration. These were then correlated over all subjects for an estimate of test-retest reliability. An Analysis of Variance was used with type of scenario (good or poor) as the grouping variable and global Equal Opportunity Climate as the dependent variable. From the data obtained on the second administration, factor scores for each of the six factors were computed for every subject within each scenario condition. A Multivariate Analysis of Variance was performed using the factor scores as the dependent variables and scenario (good or poor) to group the subjects.

Using a scree line approach, six factors were retained from the Principal Components Analysis. These factors together accounted for 65% of the total variance with eigenvalues of 33.68 (47.4%), 3.67 (5.2%), 2.77 (3.9%), 2.16 (3.0%), 2.01 (2.8%), and 1.83 (2.6%), respectively. The factors were named:

Factor I: Overall Concern with Equal Opportunity Issues. Items deal with both race and sex discrimination, as well as administrative reactions to sexual harassment. The focus is on-base behavior.

Factor II: Differential Behavior by Commanders. Do commanders treat minorities differently.

Factor III: Stereotypes. These items deal with stereotypic treatment of minorities and women (e.g., females being mistaken for secretaries).

Factor IV: Sex Role Definition. These items imply that the military is a man's job.

Factor V: Overt Sexual Harassment. Items deal with superiors' demands for sexual favors from subordinates.

Factor VI: Covert Sexual Harassment. These suggest women should be decorative and subordinate to men.

In the first administration, MISEOC exhibited a high degree of internal consistency. The Cronbach's alpha over all items was .98, and for two random halves .96 and .97, respectively. The six scales were also highly reliable (alphas of .95, .93, .91, .87, .86, .88, respectively). The correlation between the two halves was .88. On the second administration, the reliability of the six scales was satisfactory (.92, .93, .86, .84, .89, .87, respectively). When the target was changed to the constructed locales, the reliability of the scales remained quite good (.94, .93, .91, .82, .85, and .76 respectively).

The good and poor scenarios produced the desired effects. The global judgment means for the two scenarios were significantly different (F=63.07, df=1, 59, p<.00001) and in the expected direction [Mean (good)= 3.46, Mean (poor)=2.62]. The MANOVA used type of locale as the independent variable and scores on the six scales as criteria. The Multivariate F was highly significant (Mult F=3.46, df=12, 104, p <.0001) as were the univariate tests for five of the six scales. Overt Sexual Harassment (Scale V) was non-significant. The means were all in the expected directions (Figure 2).

Discussion

The results suggest MISEOC has both high reliability and construct validity. It appears to be sensitive to changes in the external world. Thus, the beginning stages of construction and validation of an instrument to assess Equal Opportunity Climate have been completed. We have every reason to believe the technique will prove to be both reliable and valid when tested in the field. The next phase of validation will require more piloting at DEOMI and testing pilot versions in the field. Further reliability checks and cross validation of the factor structure should be undertaken. After these steps have been completed and the instrument revised accordingly, field validation can begin. The full field validation should use at least 20 military sites (four from each service, two in CONUS and two outside of CONUS). The instrument should be administered to a large number of subjects and convergent validation checks made using objective management indices.
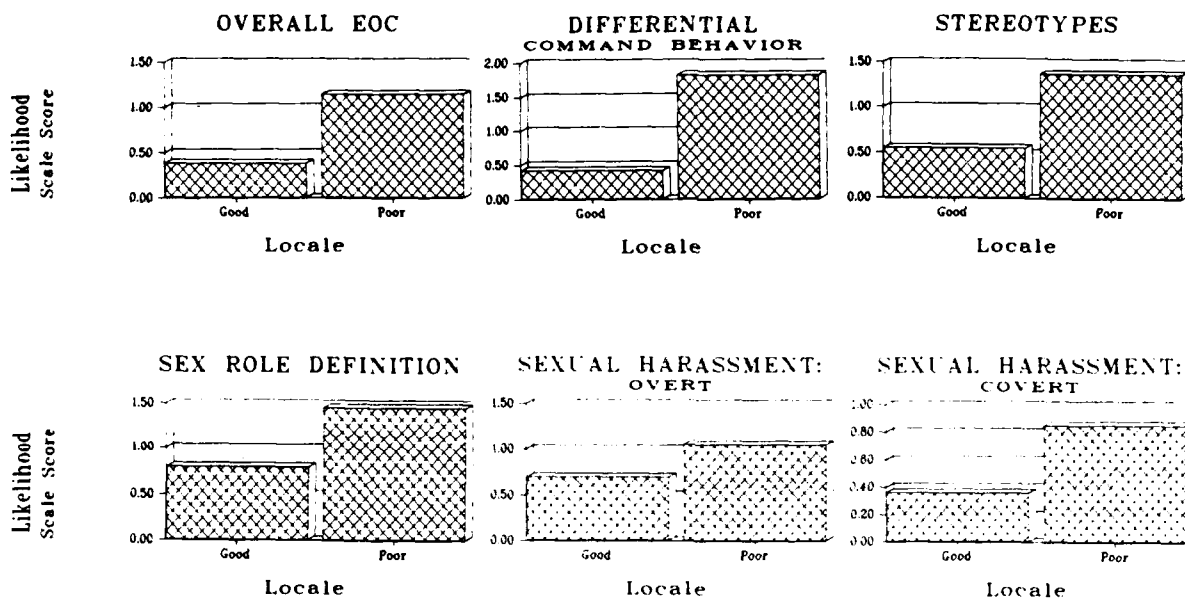
490

OVERALL EOC   DIFFERENTIAL COMMAND BEHAVIOR   STEREOTYPES

SEX ROLE DEFINITION   SEXUAL HARASSMENT: OVERT   SEXUAL HARASSMENT: COVERT

Figure 2. Effect of Locale on Equal Opportunity Climate Scale Scores.

References

Brown, D. K., Nordlie, P. G. & Thomas J. A. (1977). Changes in black and white perceptions of the Army's race relations/equal opportunity programs--1972 to 1974. Technical Report #TR-77-B3. Alexandria VA: U.S. Army Research Institute.

Bowers, D. G. (1975). Navy manpower: Values, practices, and human resources requirements. University of Michigan, Institute for Social Research.

Department of the Air Force. (1985). Air Force Human Relations Climate Assessment. Air Force Pamphlet 30-13.

Hiett, R. L., & Nordlie, P. G. (1978). An analysis of the unit race relations training program in the U. S. Army. Alexandria, VA.: U. S. Army Research Institute.

Mowday, R. T., Steers, R. M., & Porter, L. W. (1979). The measurement of organizational commitment. Journal of Vocational Behavior, 14, 224-247.

Parker, W. S. (1974). Differences in organizational practices and preferences in the Navy by race. University of Michigan, Institute for Social Research.

Pecorella, P. A. (1975). Predictors of race discrimination in the Navy. University of Michigan, Institute for Social Research.

Reily, P. J. (1980). Sexual harassment in the Navy. Thesis, Naval Postgraduate School, Monterey, CA.

Short, L. O. (1985). The United States Air Force Organizational Assessment Package. LMDC-TR-85-2. Maxwell AFB, AL: Leadership and Management Development Center, Air University.

A Comparison of Male-Female Job Proficiency
Among First-Term Navy Radioman (RM) Personnel

Herbert George Baker, Ph.D.
Virginia M. Berry, Ph.D
Navy Personnel Research and Development Center

Summary

In response to Congressional and Department of Defense (DoD) concerns, the Services are investigating measurement approaches that might be used to make personnel classification more performance based. A previous report (Laabs & Berry, 1987) explicated the research strategy of the Joint-Service Job Performance Measurement/Enlistment Standards Project. Basically, that strategy is to construct hands-on performance measures and investigate the use of these measures as criteria for predictor validation. A major research focus is on the development of job sample tests of technical proficiency to be used as a high-fidelity benchmark measure against which surrogate, or substitute, measures will be compared.

The global problem addressed by this research is the development of measurement technologies that will facilitate the task of relating enlistment standards to actual job performance. Within that overall effort, it is necessary to develop performance measures for the Navy Radioman (RM) rating. Because (1) a major career point is reached by survival to the end of the first enlistment, and (2) the emphasis is on apprentice level job performance prediction, the effort reported herein is focused on first-termers (i.e., job incumbents with 48 months or less service).

The RM rating is one of the jobs selected by the Navy for performance test development because (1) it is critical to mission success; (2) it has a large population, including substantial numbers of women and ethnic minorities; (3) it includes assignments on ships and at shore facilities; and (4) it is similar to RM jobs in the other Armed Services (Laabs, Baker, Kroeker, & Kidder, 1986). A number of performance measures were developed for this rating including (1) a hands-on job sample test, (2) a job knowledge simulation test, and (3) a set of rating scales.

Data were collected for a large sample of RMs at three locations. The sample included shipboard and shore-based personnel, from large and small installations. Data analyses were conducted to compare the performance of male and female first-term RMs on the hands-on test and their evaluations on the peer- and supervisor-level rating scales.

Stressors of Senior USAFA Female Cadets who are Non-athletes and on Academic Probation

David D. Baldessari,
Craig D. Goodrich,
David E. Peters,
Benjamin R. Sanders
United States Air Force Academy

Abstract

The primary stressors of a target group (non-intercollegiate, female, senior cadets on academic probation at the U.S. Air Force Academy) were investigated and compared with a random sample of other USAFA cadets. Identical surveys were given to members of both the control and target groups and each were asked to rate thirty different stressors on a ten point, Likert-type scale. The control group consisted of thirty first class cadets picked randomly throughout four squadrons, and included four females and twenty-six males. The results of the study showed that the target group experience significantly more stress in the areas of privilege losses and probability of dismissal. The target group also experienced significantly less stress over roles and relationships in squadrons. This study proposes that counseling be given to the target group and that the restrictions that come with academic probation should be made less severe. This would lower overall stress and give more responsibility and autonomy to the people on academic probation.

THE PILOT'S INSTRUMENT CROSSCHECK:
EFFECTS OF WORKLOAD DURING SKILL ACQUISITION

Joseph L. Bunecke
U.S. Air Force Academy

This study quantitatively demonstrates the information processing demands of one aspect of a pilot's instrument scan, or crosscheck. When interpreted in the context of a generic flight training syllabus, the results indicate that the workload student pilots experience during training significantly affects the quality of the perceptual-motor skills they eventually develop.

In a general sense, two tasks confront a student pilot learning instrument flight procedures: basic aircraft control (i.e., the moment-to-moment control of the altitude, heading, and velocity vectors) and navigation (i.e., maintaining a specific groundtrack). The control-performance technique of visual scanning dictates establishing initial parameters on the control instrument (i.e., the attitude indicator), and monitoring the performance instruments (i.e., altimeter and heading and airspeed indicators) to insure that actual performance matches desired performance. To assess the resource demands of using a control-performance crosscheck to maintain basic aircraft control, this experimental paradigm compares sequential instrument presentation (where presentation rate and order correspond to the cause-effect relationship between instruments) to free access presentation (where rate and order are subject determined). When subjects demonstrated proficiency, those forced to interpret the instruments sequentially were significantly slower than those in the free access group.

Assuming that basic aircraft control is effortful and that navigation consumes the majority of attention resources, the crosscheck might reflect the student's attempts to minimize workload. Since instructors typically evaluate progress on the basis of aircraft performance, it's not surprising that task-saturated students adopt a scanning strategy which fixates on the performance instruments. Such a stagnating crosscheck decreases net resource demand by eliminating the need to formulate the pitch/altitude, bank/heading, and pitch/power/airspeed combinations. This maladaptive scan pattern does not exhibit the natural dependency between the instruments and, hypothetically, might eventually become ingrained and used in any high workload situation. In the final analysis, I feel that the introduction of navigation principles before the information processing of a control-performance crosscheck becomes automatic, as a factor contributing to the formation of the stagnating/fixating visual scans implicated as causes of some aircraft accidents.

The Seven Alameda Health Practices
and Physical Fitness among Navy Shipboard Men[1]

Linda J. Dutton
Terry L. Conway

Health Psychology Department
Naval Health Research Center
P. O Box 85122
San Diego, CA 92138-9174

The Seven Alameda Health Practices
and Physical Fitness among Navy Shipboard Men

Linda J. Dutton and Terry L. Conway
Naval Health Research Center

The Alameda County survey of 1965 and several follow-up assessments established that seven common health practices--adequate sleep, regular breakfast, physical activity, avoiding between-meal snacks, moderation in alcohol consumption, weight control, and not smoking--were associated positively with good health, inversely with mortality. These surveys did not include active duty military personnel, however. This report examines the relationship between health practices and physical fitness in a group of Navy men. The purpose of the present study was fourfold: (a) to describe prevalence estimates for the seven Alameda health habits among shipboard men; (b) to compare these estimates to those for men nationwide as well as for male Air Force personnel; (c) to outline the practice of health habits by age, race, and education; and (d) to analyze the relationship between good health practices and scores on several tests of physical fitness.

Method

Data were collected from 1357 men stationed aboard nine Navy ships. Participants ranged in age from 18 to 51 years (mean age 26 years). The men completed a self-report questionnaire concerning their lifestyle habits, including the Alameda health practices listed above. They were also evaluated on a four-part test of physical fitness as mandated by Navy regulations, including: (a) time on a 1.5-mile run, (b) number of bent-knee sit-ups completed in two minutes, (c) inches reached on a sit-reach flexibility test, and (d) percent body fat. In addition to raw scores on these four tests, an Overall Fitness score was computed for each individual, using the mean of the standardized z scores for each test. A "total health" score was also calculated for each person by summing their good health habits, allowing one point for each good habit.

Results and Discussion

Compared to men nationwide, the total health scores of Navy men were lower. Navy men reported sleeping less, smoking more, and drinking more than U.S. civilian men. However, they also reported exercising more and snacking less often. Compared to the Air Force sample, more Navy men reported smoking and drinking. Demographic differences among the Navy sample indicated that Blacks have better health habits than Whites, and men with more education have better habits than those with less; age differences were inconsistent, and age was uncorrelated with total health score. Analysis of variance revealed significant differences among groups (based on total health score) for all fitness measures: men who engaged in positive health behaviors had higher mean fitness test scores than men who practiced fewer good health habits. The Navy's comprehensive Health and Physical Readiness Program is aimed at improving health practices in these areas and should produce a healthier and fitter naval force.

THE USE OF AN EYE TRACKING DEVICE FOR THE
MEASUREMENT OF FLIGHT PERFORMANCE
IN SIMULATORS

Lt Kevin W. Dixon
Lt Victoria A. Rojas

ABSTRACT

Design and evaluations of simulator visual systems have historically depended on subjective questionnaires and objective pilot performance data to determine training effectiveness. These types of data give the researher an idea of pilot impression and performance, but do not allow objective determinations of the pilots visual activity. This problem can be addressed in part, by the use of an eye-tracker that will be described in this paper. The eye monitoring system will provide a new set of data that can be analyzed in conjunction with or separate from traditional types of data. The eye-tracker produces data on the eye focal point which allows the researcher to determine what the pilot is focusing on at any time during the mission profile. The eye-tracker does not inhibit head movement, is relatively light weight, and is easy to calibrate and use. The elements in a fully configured system include a camera (attached to a headband worn by pilot), photoelectric sensors, an eye monitor, a VCR tape recorder, a time code generator, and an offline personal computer. The system was initially used in a study to determine field of view requirements for the C-130 weapon system trainer. This system showed that pilot eye movements can be tracked in a simulator environment to allow field of view recommendations. Other potential visual training research applications include determinations of visual workload, analyses of real-time attention allocation, comparisons of pilots crosschecks, decisions on scene content for low level flight, and target detection in the visual field.

Microcomputers in the Dormitories (MIDs):
A Comparison of USAF Academy Cadet Attitudes and Use

Gail A. Fallon, M.S.
Timothy E. Evans
United States Air Force Academy

Abstract

The data reported in this study is part of ongoing research into the impact of requiring all cadets, starting with the class of 1990, to have microcomputers and access to a Local Area Network (LAN) in the dormitories. All of the class of 1990 were given 20 bipolar semantic differential, 7-point rating scales to assess their general attitudes about computers and a questionnaire assessing their experience level prior to entry to the USAF Academy. Fifteen freshmen were then compared to fifteen juniors to see if there were any differences in attitudes and experience between the two classes. The freshmen as a class were more positive in their ratings on all 20 of the items assessing general attitude towards computers, with half of the attitude items being significantly different ($p<.05$). A factor analysis of the attitude survey revealed that the two classes were most likely to differ on items with high loadings on factor I, expressing human qualities (e.g., good/bad, humane/inhumane) and on items with high loadings on factor V, expressing ease of use (e.g., complicated/uncomplicate, overpowering/easy to control). The two classes were least likely to differ on items with high loadings on factor II, expressing more utilitarian items (e.g., efficient/inefficient, time saving/time consuming, costly/economical). The differences in attitudes between the classes does not seem to be due to differences in computer experience since no significant differences were found in the type of computing experience (e.g., games or wordprocessing) or the place where one got one's experience (e.g., school or at home).

Interpreting Maps with Plan-View vs. Perspective Geometry

Ronald D. Lyons, Jeffrey M. Powers, & Marilyn Vala-Rossi
Human Factors Engineering/System Safety
General Dynamics Land Systems
Dean G. Purcell & Frank Schieber
Department of Psychology
Oakland University

Abstract

Training experience with U.S. Army recruits has revealed a wide range of spatial reasoning abilities across individuals. These differences in spatial abilities present themselves in terms of specific skills such as pathfinding and target acquisition/avoidance, using detailed contour terrain maps. Recruits with poor spatial skills often require unacceptablly long intervals to extract critical navigation and targeting information from contour maps. Under conditions of stress (e.g., combat) map reading speed and accuracy among these personnel can be expected to deteriorate even further. The use of electronic imaging displays to convey contour-terrain map information to tank crews offers several opportunities to lessen many difficulties encountered in the transfer of spatial information. The present study was conducted to explore the potential utility of monochrome and color-coded 3D perspective maps for conveying altitude information during the rapid performance of cartographic tasks.

The utilization of three different types of computer generated graphic contour maps was investigated in this study: monochrome plan-view, monochrome 3-D, and color enhanced 3-D maps. Twenty-four male volunteers engaged in multiple tasks with one of the map types. Subjects in each Map-Type groups performed three different map interpretation tasks: determining if there was an obstruction of view between two sites presented on a map, determining if a second target was higher or lower in altitude relative to a previously presented target superimposed on the map, and tracing the easiest route of travel.

The results indicate that 3-D maps have advantages over plan view maps, and that this advantage is amplified when color is added. The color perspective view map was associated with lower reaction times and errors than the monochrome perspective view and plan view maps. Potential benefits of such perspective displays include reducing the latency of route-planning and greater efficiency and accuracy in the utilization of the z-axis (altitude) in tactical maneuvering.

How to Use a Mouse to Simulate Learning
Mary Z. Mays and Charles L. Winstead, Jr.
U.S. Air Force Academy

A computer simulation of learning was developed for use with the
Macintosh Plus[TM] computer. the underlying model is derived from those
of Rescorla and Wagner (1972) and Daly and Daly (1982; 1984). The
simulation allows a student to conduct an experiment in learning.
Changes in learning over trials are assumed to be the result of the
subject's previous experience in the experimental setting, the
subject's motivation, the magnitude of reinforcement received for a
response, and the saliency of cues present in the setting. The
student uses the Macintosh's[TM] mouse to select values for these
variables from among the options presented. Predicted changes in
learning over trials are computed and printed in a few minutes. Thus,
students can quickly replicate classic learning experiments, compare
the predictions of different theories, or test their own hypotheses
about learning.

Daly, H. B., Daly, J. T. (1982). A mathematical model of reward and
    aversive nonreward: Its application in over 30 appetitive situations.
    Journal of Experimental Psychology: General, 111, 441-480.

Daly, H. B., Daly, J. T. (1984). DMOD--A mathematical model of reward and
    aversive nonreward in appetitive learning situations: Program and
    instruction manual. Behavior Research Methods, Instruments, and
    Computers, 16, 38-52.

Rescorla, R. A., Wagner, A. R. (1972). A theory of Pavlovian
    conditioning: Variation in the effectiveness of reinforcement and
    nonreinforcement. In A. H. Black, W. F. Prokasy (Eds.), Classical
    conditioning II: Current research and theory (pp. 64-99). New York:
    Appleton-Century-Crofts.

Using Portable Microcomputers to Conduct Organizational Surveys[1]

Paul Rosenfeld, Ph.D.
Linda M. Doherty, Ph.D.
S. Mitchell Vicino
Jean Greaves
Navy Personnel Research and Development Center

Abstract

Surveys are a useful means of obtaining attitudinal information in both military and civilian organizations. During the past several years, researchers at the Navy Personnel Research and Development Center in San Diego, CA have developed and tested several versions of a computerized survey system as part of the CENSUS (Computerized Executive Networking Survey System) project. Using a networked version of CENSUS, a computer survey has been administered to up to sixteen individuals simultaneously using an IBM/AT host and Qume terminals. An individualized version called MASQ (Microcomputer-based Assessment, Surveys, and Questionnaires), allows a survey to be administered on any standard IBM or IBM compatible microcomputer. Studies using both the networked and individualized systems have successfully reduced error rate and increased user enjoyment. However, because organizations may not be able to provide the necessary computer equipment or space to run CENSUS or MASQ, it would useful if a portable system were available whereby the survey team could themselves provide the computers used for the survey. Such a system, P-MASQ (Portable MASQ), has recently been successfully tested in a survey conducted with several hundred undergraduate college students at San Diego State University. In that project, a computer survey was administered on three Zenith-181 and one IBM-convertible portable computers. The results showed that the responses obtained on a number of psychological and organizational items were virtually identical to those obtained on comparable paper-and-pencil surveys and comparable to previously gathered data on high resolution PC computer monitors. Furthermore, user satisfaction and enjoyment were high. P-MASQ retains the numerous advantages of computerized surveys (e.g., speed, reduction of missing responses, facilitation of data analyses) while adding the portability feature. Thus, it is now possible to conduct computer surveys in organizations without requiring large allocations of resources or space. Also, P-MASQ allows computer surveys to be conducted at remote work sites, thus alleviating the need for employees to leave their work place and lose time on the job by reporting to a central survey site.

---

[1] The views expressed are those of the authors, they are not official and do not necessarily represent the views of the Navy Department.

**APPLICATION OF JOB RESOURCE FRAMEWORK
TO FLIGHT CREW MANAGEMENT**

POSTER SESSION

LT COL FRANK WOOD, et.al.

Military Psychology: A Journal of the Division of Military Psychology
American Psychological Association

Martin F. Wiskoff, PhD.
Defense Personnel Security Research and Education Center
Monterey, CA

The Division of Military Psychology (Division 19) of the American Psychological Association, in conjunction with Lawrence Erlbaum Associates, Inc., plans to publish a new quarterly journal starting in January 1989. Military Psychology invites the submission of manuscripts in accordance with the following guidelines for contributors.

CONTENT: Military Psychology is the quarterly journal of the Division of Military Psychology (Division 19) of the American Psychological Association. The journal seeks to facilitate the scientific development of military psychology by encouraging communication among researchers and practitioners.

The domain of military psychology is the conduct of research or practice of psychological principles within a military environment. The journal publishes behavioral science research papers having military applications in the areas of clinical and health psychology; cognition and training; human factors; manpower and personnel; social and organizational systems; and testing and measurement. One issue per year will be devoted to comprehensive treatment of a single topic of major concern to military psychology.

Military Psychology is international in scope, and the editors encourage submission of articles that address appropriate research being carried out in a variety of national settings.

Contributions will be considered for publication in the following categories:

Research Articles...reports of empirical research.
Notes...brief papers describing work that is largely confirmatory, advances in knowledge arising as by-products of broader studies, or new research techniques and methodologies.
Reviews...scholarly integrations of individual areas of empirical research.
Communications...information on policies and trends that affect the support and direction of research in military psychology.

Comments and letters to the editor will be published on a space-available basis.

SUBMISSION OF MANUSCRIPT: Submit five copies of the manuscript to Martin F. Wiskoff, Editor, Military Psychology, Defense Personnel Security Research and Education Center, 99 Pacific Street, Suite 455-E, Monterey, CA 93940.

COVER LETTER: The cover letter should request publication of the manuscript in Military Psychology and should include (a) a statement that the manuscript is not being considered simultaneously for publication elsewhere; (b) confirmation that all original research procedures were consistent with the principles of research ethics published by the American Psychological Association, except as may be detailed in the manuscript; and (c) a request for blind review, if desired.

MANUSCRIPT PREPARATION: Manuscripts should be double-spaced, highly legible and prepared according to the Publication Manual of the American Psychological Association, Third Edition, which may be obtained from APA, 1200 Seventeenth Street, NW, Washington, DC 20036. Manuscripts will not be returned to authors.

Research manuscripts should be no more than 20-25 pages, including tables, figures, and references. The title page should specify the manuscript title, names and institutional affiliations of all authors exactly as they are to be printed, name and addresses of the author to whom requests for reprints should be addressed, and a running head of no more than 60 letters and spaces. Where blind review has been requested, all identifying information including footnotes of acknowledgement should be confined to the title page. The second page should include the manuscript title and an abstract of 100-150 words. Supply black and white originals or camera-ready glossy reproductions of all illustrations.

MANUSCRIPT REVIEW: Manuscripts will be evaluated on the basis of style as well as content. Except for minor copyediting changes, authors will be responsible for clarity, conciseness, and felicity of expression. Authors will be responsible for proofreading copyedited manuscripts and page proofs and returning them promptly to the Editor. Reprints may be ordered at the time page proofs are returned.

PERMISSIONS: Contributors are responsible for all statements made in their work and for obtaining permission from copyright owners if they use an illustration or lengthy quote (over 100 words) published elsewhere. Contributors should write to both publisher and author of such material, requesting nonexclusive world rights in all languages for use in the present article and in all future editions of it.

AUTHOR INDEX